ucc

Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

Investigation of Glycopolymer Assembly Systems in *Lactococcus lactis*

A Thesis Presented to the National University of Ireland Cork

By

Ilias Theodorou

MSc Biotechnology

For the Degree of

PhD in Microbiology

School of Microbiology

National University of Ireland Cork, June 2019

Supervisors: Prof. Douwe van Sinderen and Dr. Jennifer Mahony

Head of the Department: Prof Gerald Fitzgerald

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

This work is dedicated to my greatest supporter and guide

My mother

Table of Contents

Abbreviations 4

Abstract 7

Chapter 1 Introduction 11

Chapter 2 Identification of genes involved in glucose decoration of the *Lactococcus lactis* cell wall polysaccharide 65

Chapter 3 Modification of lactococcal cell wall-associated rhamnan and lipoteichoic acids and associated impact on phage infection 107

Chapter 4 A dual chain assembly mechanism to synthesize a complex cell wall rhamnopolysaccharide with high structural diversity in *Lactococcus lactis* 155

Chapter 5 Isolation and characterization of bacteriophage escape mutants infecting PSP-deficient *Lactococcus lactis* derivatives 210

Chapter 6 Discussion 237

Acknowledgements 244

Abbreviations

aa = Amino Acid

ABC = ATP binding cassette

BIM = Bacteriophage insensitive mutant

BppA = Accessory baseplate protein

BppL = Lower baseplate protein

BppU = Upper baseplate protein

C_{55}-P = Undecaprenyl phosphate

CBM = Carbohydrate binding module

CDS = Coding sequence

Cm^r = Chloramphenicol resistance

CPS = Capsular polysaccharide

CRISPR = Clustered regularly interspaced short palindromic repeats

CWPS = Cell wall associated polysaccharide

D-ala = D-alanine ester

DAG = Diacylglycerol

DHB = 2,5-dihydroxy-benzoic acid

Dit = Distal tail protein

DMSO = Dimethylsulfoxide

dTDP = Deoxythimidine diphosphate

dTTP = Deoxythymidine triphosphate

EPS = Exopolysaccharide

Ery^r = Erythromycin resistance

FID = Flame ionization detector

Gal = Galactose

Gal*f* = Galactofuranose

GalNAc = N-acetyl-galactosamine

GC – MS = Gas chromatography mass spectrometry

GC – FID = Gas chromatography flame ionization detector

Glc = Glucose

GlcNAc = N-acetyl-glucosamine

GRAS = Generally regarded as safe

Gro = Glycerol

GroP = Glycerolphosphate

GT = Glycosyltransferase

HF = Hydrofluoric acid

HPAEC = High performance anion exchange chromatography

Ino = m-Inositol

LAB = Lactic acid bacteria

LCP = LytR-CpsA-Psr protein family

LPS = Lipopolysaccharide

LTA = Lipoteichoic acid

MALDI-TOF MS = Matrix-assisted laser desorption ionization-time of flight mass spectrometry

ManNAc = N-acetyl-mannosamine

MS = Mass spectrometry

m.u. = Mass unit

MurNAc = N-acetyl-muramic acid

NICE = Nisin-controlled gene expression

NMR = Nuclear magnetic resonance

OD = Optical density

ORF = Open reading frame

ONPG = Ortho-nitrophenyl-β-galactoside

PAD = Pulse-amperometric detection

PAM = Proto-spacer adjacent motif

Pip = Phage infection protein

PG = Peptidoglycan

PolM = Polytopic transmembrane protein

PS = Polysaccharide

PSP = Polysaccharide pellicle

RboP = Ribitolphosphate

RBP = Receptor binding protein

Rha = Rhamnose

RT – PCR = Reverse transcription PCR

SDS = Sodium dodecyl sulfate

SEC-HPLC = Size exclusion chromatography – high performance liquid chromatography

SNP = Single nucleotide polymorphism

ssDNA = Single stranded DNA

TA = Teichoic acid

Tal = Tail associated lysin

Tet^r = Tetracycline resistance

TFA = Trifluoroacetic acid

TMH = Transmembrane helices

TSS = Transcription start site

TU = Transcriptional unit

UndP = Undecaprenyl phosphate

WTA = Wall teichoic acid

Abstract

The lactic acid bacterium *Lactococcus lactis* is the most widely employed species in dairy fermentations, and due to this economic significance extensive research pertaining to its functionality, physiology and interaction with its environment has been performed (1). Since the discovery of their disruptive impact on fermentations, extensive scientific efforts have been directed to elucidate the molecular interactions between *L. lactis* and its infecting bacteriophages (phages). The lactococcal cell wall consists of a thick peptidoglycan layer as well as its associated secondary polymers. It represents a formidable barrier for phages, and for this reason deserves thorough and focused investigation. In particular, several components of the cell wall have either been conclusively determined to act as receptors for phage recognition and attachment during infection, or are at least suspected to play a role in this process. In particular, polysaccharides associated with the cell wall play a crucial role in the recognition/attachment process of many phages. In this thesis, the biosynthesis of such glycopolymers as well as their contribution to phage-host interactions will be extensively discussed.

A major component of the *L. lactis* cell wall is the cell wall-associated polysaccharide (CWPS), which consists of a rhamnosyl polysaccharide, known as the rhamnan, and a phospho-polysaccharide (PSP), known as the pellicle. Previous work has established the importance of the so-called *cwps* gene cluster for the assembly of this glycopolymer (2) as well as its importance in establishing phage/host interactions (3). Following an extensive investigation of the *L. lactis* genome, seven additional genes were identified based on their homology to previously identified genes in Gram positive bacteria, encoding enzymes involved in glycopolymer assembly. Of these seven genes, six are found as three pairs, termed *csdAB*, *csdCD*, and *csdEF*, in the lactococcal genome. The first gene of each pair (*csdA*, *csdC*, and

csdE) is presumed to encode an undecaprenyl-phosphate (C_{55}-P) activating glycosyltransferase (C_{55}-P GT), while the second gene (*csdB*, *csdC*, and *csdE*) encodes a putative polytopic glycosyltransferase (PolM GT). Lastly, a flippase-encoding gene, termed *cflA*, was also identified. Each pair of the GT-encoding genes along with the flippase-encoding gene could putatively encode for a three-component glycosylation system previously encountered in several Gram-negative and Gram-positive bacteria (4).

The specific function of each of the three *csd* gene pairs was investigated through a mutational approach, whereby a non-sense mutation was introduced into the above-mentioned genes either individually or in combination using single stranded DNA (ssDNA) recombineering. The effect of these mutations on various saccharidic structures associated with the lactococcal cell wall was investigated by means of high performance anion exchange chromatography (HPAEC), matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS), and gas-chromatography mass-spectrometry (GC-MS). Our analysis was successful in determining the functionality of each of the three gene pairs (*csdAB*, *csdCD*, *csdEF*). More specifically, CsdA/CsdB are required for the glucosylation of the rhamnan component of the CWPS, while the CsdC/CsdD proteins are involved in PSP glucosylation. The products of *csdEF* do not appear to alter the structure of CWPS, yet were shown to galactosylate lipoteichoic acid (LTA) moieties. We present evidence that the presumed flippase, CflA, is responsible for rhamnan and, though partially, PSP glucosylation. Finally, glucosylation of lactococcal rhamnan and PSP, as imparted by CsdAB and CsdCD respectively, was shown to have an impact on phage/host interactions. No such association was identified for the galactosylation of LTA by CsdEF, however, preliminary evidence for the importance of LTA glycosylation in the cell's intrinsic bacteriocin sensitivity was observed.

A similar mutational strategy to that described above was employed for the functional genomic dissection of the *cwps* gene cluster. An extensive structural analysis of the extracted

CWPS from these individual *cwps* mutants provided insights into the various functions of the enzymes encoded by the mutated *cwps* genes. From the structural analysis of the PSP, it was shown that certain mutations exerted a more detrimental impact on PSP production than others: mutations in *wpsJ*, *wpsA*, and *wpsC* appear to cause complete elimination of PSP production, while mutations in the remaining genes, (*wpsBDEFHI*) still allowed for the production of reduced levels of (sometimes incomplete) PSP. This difference in PSP production (i.e. no or reduced levels of PSP) of these *cwps* mutants also had an impact on their phage sensitivity profiles. More specifically, it was observed that all *cwps* mutants were resistant to infection by phages that are able to infect the wild type host *L. lactis* NZ9000. However, upon sustained exposure of certain *cwps* mutants to such phages, escape mutants could be isolated that were able to infect *cwps* mutants producing reduced or incomplete PSP, whereas no such escape mutants could be obtained against strains carrying *cwps* mutations that cause complete elimination of PSP production. In tandem with the functional characterization of *cwps* genes, the transcriptional organisation of the cluster was investigated through promoter mapping and primer extension analysis. These results confirmed previous speculation on the division of the gene cluster into three distinct transcriptional units in which the 5'-end operon is dedicated to rhamnan assembly and the 3'-end operon is dedicated to PSP assembly.

The work described in this thesis has identified three distinct glycosylation systems which are variably present within members of the *L. lactis* species, and which are employed for the modification of different cell wall-associated glycopolymers. Together with the extensive characterisation of the *cwps* gene cluster, it has provided further insight into the discrete biosynthetic steps required for the assembly of complex glycopolymers such as the CWPS and LTA in *L. lactis*. Although we obtained a working model on the mode of assembly of these structures, the importance of their observed structural diversity on bacterial phenotypic differences has been far from resolved and further investigations will be required to understand these relationships and associated biological significance.

References

1. Song AAL, In LLA, Lim SHE, Rahim RA. 2017. A review on *Lactococcus lactis*: from food to factory. Microbial Cell Factories 16:15.

2. Dupont K, Janzen T, Vogensen FK, Josephsen J, Stuer-Lauridsen B. 2004. Identification of *Lactococcus lactis* genes required for bacteriophage adsorption. Applied and Environmental Microbiology 70:5825-32.

3. Mahony J, Kot W, Murphy J, Ainsworth S, Neve H, Hansen LH, Heller KJ, Sørensen SJ, Hammer K, Cambillau C, Vogensen FK, van Sinderen D. 2013. Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. Applied and Environmental Microbiology 79:4385-4392.

4. Mann E, Whitfield C. 2016. A widespread three-component mechanism for the periplasmic modification of bacterial glycoconjugates. Canadian Journal of Chemistry 94:883-893.

Chapter 1

Introduction

Table of Contents

1.1. Lactic Acid Bacteria 13

1.1.1 General Characteristics and Industrial Significance 13

1.1.2 *Lactococcus lactis* 14

1.1.3 *Lactococcus lactis* and the Bacteriophage Challenge 15

1.2. Lactococcal bacteriophages 15

1.2.1 Taxonomic classification 15

1.2.2 Phage-host molecular machinery 17

1.2.3 Gram-positive bacteriophage receptors 22

1.3. Lactococcal cell wall glycopolymers 24

1.3.1 Rhamnosyl cell wall-associated polysaccharides 24

1.3.2 Biosynthesis and assembly of the CWPS 30

1.4. Lactococcal teichoic acids 33

1.4.1 Wall-teichoic acid structure and assembly 33

1.4.2 Lipo-teichoic acid structure and assembly 38

1.4.3 Modification of teichoic acids 40

1.5. Genomic modification of lactoccocal strains 44

1.6. Conclusions and outline of this thesis 46

1.7. References 47

1.1. Lactic Acid Bacteria

1.1.1 General Characteristics and Industrial Significance

Lactic acid bacteria (LAB) comprise an order of bacterial genera, also known as *Lactobacillales*, whose main characteristics include, among others, their ability to ferment hexose sugars and release lactic acid as a primary metabolic end-product. LAB are furthermore characterised as Gram-positive, microaerophilic bacteria, with genomes of generally low GC content (1). The order of LAB includes a number of genera including *Lactococcus*, *Streptococcus*, *Lactobacillus*, *Enterococcus*, *Oenococcus*, *Leuconostoc* and *Pediococcus* (1). Representative species from these genera can be organised into two separate groups based on the specific metabolic pathway they employ. The homofermentative group converts glucose primarily into lactic acid, while the heterofermentative group produces lactic acid, carbon dioxide (CO_2), as well as ethanol and/or acetic acid (1, 2). In general, the acidification activity of LAB combined with, in some cases, the ability to produce antimicrobial compounds has underpinned their extensive application as starter cultures in the preservation of foodstuffs. They are the main driving force of microbial bioconversions that take place during the fermentation of a number of food products, including dairy, meats, and vegetables (2). Their presence in food fermentations has also been shown to contribute positively to both the organoleptic properties as well as the nutritional value of the final fermented product (2-4). The importance of LAB to the modern industrial food fermentation sector is illustrated by the multi-billion dollar value of the dairy market (5). The most widely used LAB dairy starter species include *Lactococcus lactis*, *Streptococcus thermophilus* and *Lactobacillus delbrueckii* ssp. *bulgaricus* (3, 5).

Due to the long and safe history of use of LAB in food fermentation, a large number of representative species have been granted Generally Regarded As Safe (GRAS) status, which has also led to a considerable breadth of understanding of the metabolic characteristics and capabilities of these bacteria. Through this knowledge, there has been an increased interest for

the potential application of LAB in the production of value-added biochemicals and metabolites (6-8). Such compounds include flavour compounds, e.g. acetaldehyde (9), low-calorie sweeteners, e.g. sorbitol (10, 11), and vitamins, e.g. folate (12, 13). In combination with their acquired GRAS status, recent developments in the use of LAB as mucosal delivery vectors of therapeutic molecules through recombinant expression have also placed renewed interest in their bio-therapeutic potential (14, 15).

1.1.2 *Lactococcus lactis*

Although *L. lactis* is believed to have originated from a plant-based environment (16-18), it is now one of the most commonly and extensively employed bacterial starter culture species for the fermentation of dairy products, in particular cheese (18). In terms of phenotypic taxonomy, *L. lactis* is divided into four subspecies based on characteristics, such as carbohydrate utilisation, salt tolerance and resistance to a range of antibiotics (19). This species includes *L. lactis* ssp. *cremoris*, *L. lactis* ssp. *lactis* (both isolated from milk sources), *L. lactis* ssp. *hordniae* (isolated from the leafhopper *Hordnia circellata* (19, 20)), and *L. lactis* ssp. *tructae* (isolated from two different trout species (16, 19)). In addition to the phenotypic categorisation of isolated *L. lactis* strains, there has also been considerable investigation into the genetic basis of this taxonomy. Based on extensive genotyping, two major genotypes of *L. lactis* have been identified, i.e. the *lactis* and *cremoris* genotypes (21, 22). Interestingly, *L. lactis* strains have been isolated whose apparent phenotypic profile does not match their assigned genotype (16, 23, 24), making taxonomic classification of the *L. lactis* species particularly challenging. In addition to its extensive use in food fermentation processes, *L. lactis* is becoming a very useful research model in the field of genetic engineering due to its relatively small genome size combined with the increasing number of genetic engineering tools developed specifically for this species (25). Perhaps the most important tool developed in *L. lactis* is the nisin-controlled gene expression (NICE) system (26), which utilizes the *nisRK* two component system to induce gene expression in a given *L. lactis* strain upon exposure to the

anti-microbial peptide known as nisin. Over the years, the NICE system has been widely used for various over-expression, metabolic engineering, and protein secretion studies of Gram-positive bacteria (18, 27-29).

1.1.3 *Lactococcus lactis* and the Bacteriophage Challenge

Traditionally, the production of fermented foods relied on the autochthonous microbiota of the fermentation environment and/or the raw ingredients. Owing to the high demand for food products of consistent and uniform quality, the modern food fermentation industry is reliant on the application of well-defined and known bacterial starter cultures, of which *L. lactis* is a major component (30, 31). This heavy reliance on certain strains of LAB creates the ideal conditions for (bacterio)phage predation (32). Phage-mediated interference of fermentation processes can lead to product inconsistencies or loss with serious negative economic consequences to the producer. As a result, extensive efforts have been expended aimed at understanding the intricacies of phage/host relationships and in developing novel approaches to mitigate the constant pressure that phages exert on such microbial food fermentation systems. *L. lactis* has been instrumental in understanding the multi-layered process by which phages interact with and infect Gram-positive microorganisms (33).

1.2. Lactococcal bacteriophages

1.2.1 Taxonomic classification

All phages that have been described to infect LAB fall within the *Caudovirales* order (34, 35) of tailed phages. The *Caudovirales* order is comprised of three families: the *Podoviridae* with short and non-contractile tails; the *Siphoviridae* with long and non-contractile tails; and the *Myoviridae* that possess long, contractile tails (34, 35). Due to the importance of *L. lactis* to the fermentation industry, phages that infect this species are some of the most widely investigated and studied (36-41). Currently, phages of *L. lactis* are classified as *Podoviridae* (two phage groups) or *Siphoviridae* (eight phage groups) members (35). A

summary of these species as well as the total number of fully sequenced representatives of each is presented in Table 1.

Table 1. Summary of isolated phage species that infect *L. lactis*

Phage Family	Species Name	Fully sequenced members
Siphoviridae	936	118
	P335	31
	c2	19
	1706	5
	949	13
	1358	1
	Q54	1
	P087	7
Podoviridae	P034	2
	KSY1	1

The most frequently encountered groups of phages in dairy fermentations are the 936 group, followed by members of the P335 and c2 groups, which collectively account for 98 % of dairy fermentation isolates (35, 42-46). Members of the 936 group are characterised by isometric heads and are strictly virulent (45), while the strictly virulent c2 phages exhibit a prolate capsid (47, 48). P335 phages possess isometric capsids and members of this group may be virulent or temperate (49, 50). The apparent prevalence of the 936, P335, and c2 group phages within the dairy industry may be a result of numerous factors including the consistent use of a limited number of lactococcal bacterial starter cultures, the ability of these phages to propagate to high titres combined with their resistance to common pasteurisation practices employed in these facilities (34, 35). Members of both the 936 (40, 51) and c2 (52) groups exhibit a high degree of genomic conservation with the highest divergence observed in the

early- and middle-expressed regions of their respective genomes, which are involved in DNA replication while certain structural-protein-encoding genes involved in host recognition are also observed to be divergent. In contrast, isolates belonging to the P335 group exhibit much greater genetic divergence and have been classified into five distinct genotypic subgroups, I-V (35, 50, 53). Genes within the genomes of P335-type phages are commonly grouped into a number of distinct functionality modules. These include genes that contribute to DNA replication functions, genes that encode for structural components, genes that are responsible for host lysis, and finally, though this applies only to the temperate members of this group, genes involved in the lysogeny life cycle (41, 50).

1.2.2 Phage-host molecular machinery

The major structural components of *Siphoviridae* phages infecting *L. lactis* include the double-stranded DNA containing capsid, the neck/collar and the tail structures. It has long been accepted that the distal part of the tail structure is responsible for the primary interaction with a molecular component of the host's cell wall, which in the case of *L. lactis*, is believed to be a saccharidic (54-57) and/or a proteinaceous compound (52).

As mentioned previously, the c2 phages exhibit the highest degree of sequence relatedness among the three most commonly isolated phage groups (52). Additionally, c2 phages have been shown to employ a two-step mode of infection whereby they first bind reversibly to a carbohydrate receptor on their lactococcal host and then attach irreversibly to a protein component of the lactococcal cell membrane (58). More specifically, it was shown that sugar monosaccharides such as rhamnose can inhibit the adsorption of c2 suggesting that the phage recognises a carbohydrate component present in or on the lactococcal cell wall, although the precise identity of this carbohydrate receptor remains elusive. In addition, two proteinaceous receptors for the irreversible binding by c2 group phages have been identified thus far in *L. lactis*, Pip (phage infection protein) and its homolog YjaE (52). Furthermore, the phage-encoded genetic components responsible for the recognition of Pip/YjaE, i.e. the

receptor binding proteins (RBP), have been identified (52). However, structural data and information regarding the assembly of the capsid and tail components and the molecular interaction of their encoded RBPs is still lacking and requires further investigation.

Unlike the c2 phages, the genetic diversity observed within the P335 phage group is considerably more pronounced to the extent that no single genetic feature is shared by all representative members (41, 50). However, based on the sequence identity of the genomic region encoding the adhesion modules, also known as the baseplates, the P335 phages have been categorised into five distinct sub-groups, designated as I – V (50, 53). Sub-group I phages, one representative of which is phage BK5-T, possess a two-component baseplate composed of an "evolved" distal tail (Dit) protein and a fused tail-associated lysin (Tal)–RBP protein that presents as a distinctive long fibre protruding from the tip of the phage's tail (Figure 1). This sub-group's Dit protein is characterised as "evolved" due to the fact that it contains a putative carbohydrate binding module (CBM), commonly absent in what are known as "classical" Dit proteins (40). Phages belonging to sub-group II all contain in their baseplate "classical" Dit proteins that are smaller than the sub-group I Dit proteins and lack a CBM while also exhibiting long Tal proteins (906 – 929 amino acids [aa]) that contain peptidoglycanase modules (41). Sub-group II baseplates are completed with an assembly of proteins known as baseplate proteins (Bpp) that include in a majority of the strains the upper Bpp (or BppU) and the lower Bpp (or BppL). BppL is the RBP of this sub-group's baseplate. A number of sub-group II representatives, including Tuc2009, also incorporate an accessory Bpp (or BppA) (Figure 1). Finally, strains belonging to sub-groups III or IV exhibit a small "stubby" distal tail structure that is composed of Dit and Tal proteins that closely resemble the equivalent proteins seen in 936-type phages, such as p2, and also incorporate relatively small RBPs (342 – 349 aa) (41) (Figure 1). Recently, a fifth sub-group has been identified whose baseplate structurally resembles that of sub-group I, containing an "evolved" Dit and a fused Tal-RBP protein, while simultaneously lacking any significant conservation in the remaining structural components of the phage (53). Significant efforts aimed at elucidating the molecular structure of the

baseplate/RBP of model P335 phages such as TP901-1 and Tuc2009 (sub-group II phages) (59-63) have been instrumental to our understanding of both the architecture of P335 adhesion modules and how these molecular structures can influence phage-host interactions.

The genomes of 936 group phages all exhibit similar transcriptional units that include the early-, middle-, and late-expressed regions with the first two together containing the DNA replication module (40, 44). Genes within the late-expressed region encode proteins responsible for functions including DNA packaging, morphogenesis and cell lysis (44). Interestingly, genetic diversity has been observed in "hot spots" associated with certain genes found in this module, including the RBP (40). The RBP of 936 phages facilitates the adsorption of the phage onto the cell surface through recognition of, and binding to, a saccharidic component known as the cell wall polysaccharide (CWPS, more on this molecular structure in later sections) (33, 64). The genetic diversity of the RBP-encoding genes corresponds to the genomic diversity of the gene clusters associated with the biosynthesis and assembly of the various CWPS polymers produced by *L. lactis* strains (40, 55). The RBPs of 936 phages belong to one of five groups (I-V) based on the sequence identity of the RBP-encoding genes. This classification is linked to the host range of the phage, i.e. phages of the same RBP-group infect bacteria belonging to the same *cwps* genotypes (40, 65). The RBPs of 936 phages have also been shown to structurally mirror those of P335 phages with recognisable head, neck, and shoulder domains (33, 66). The head domain interacts with the host-encoded receptor moiety, the shoulder domain connects the RBP with the rest of the phage tail structure, while the neck domain connects the head and shoulder domains. The RBP of the 936 group phage p2 was observed to exist in two conformations - one in which the RBP of the baseplate faces away from the host while in the other, the RBP faces towards the host and is thought to be "infection-ready". The change between the two conformations is thought to be triggered/stabilized by the presence of cationic ions, specifically Ca^{2+} (67). Interestingly, all 936 group phages tested so far require a bivalent cation, such as Ca^{2+}, to be infective, while the majority of P335 group

phages, with the exception of some sub-group IV representatives, do not exhibit such requirement (68).

1.2.3 Gram-positive bacteriophage receptors

The RBP of phages that infect Gram-positive bacteria, including those of LAB phages, have been shown to interact with several host-produced macromolecules that exist or form part of the bacterial cell wall (64, 69). A summary of these host-encoded molecules as well as the phages that recognise them is presented in Table 2. In the case of LAB, the cell wall-associated molecules that have been identified as phage receptors are either proteinaceous in nature or represent glycopolymers. As mentioned above, two proteins, Pip and YjaE, have been identified as receptors for phages belonging to the lactococcal c2 phage group (52). The *Lactobacillus delbrueckii* phage LL-H specifically recognises the poly-glycerophosphate (poly(Gro-P)) lipo-teichoic acid (LTA) (70). Finally, the cell wall polysaccharide (CWPS) of at least three *L. lactis* strains has been both structurally elucidated and shown to be involved in phage adsorption (55-57). Details on the structural characteristics as well as the genetic and molecular components involved in the assembly of these CWPS glycopolymers will be presented in the ensuing sections of this review.

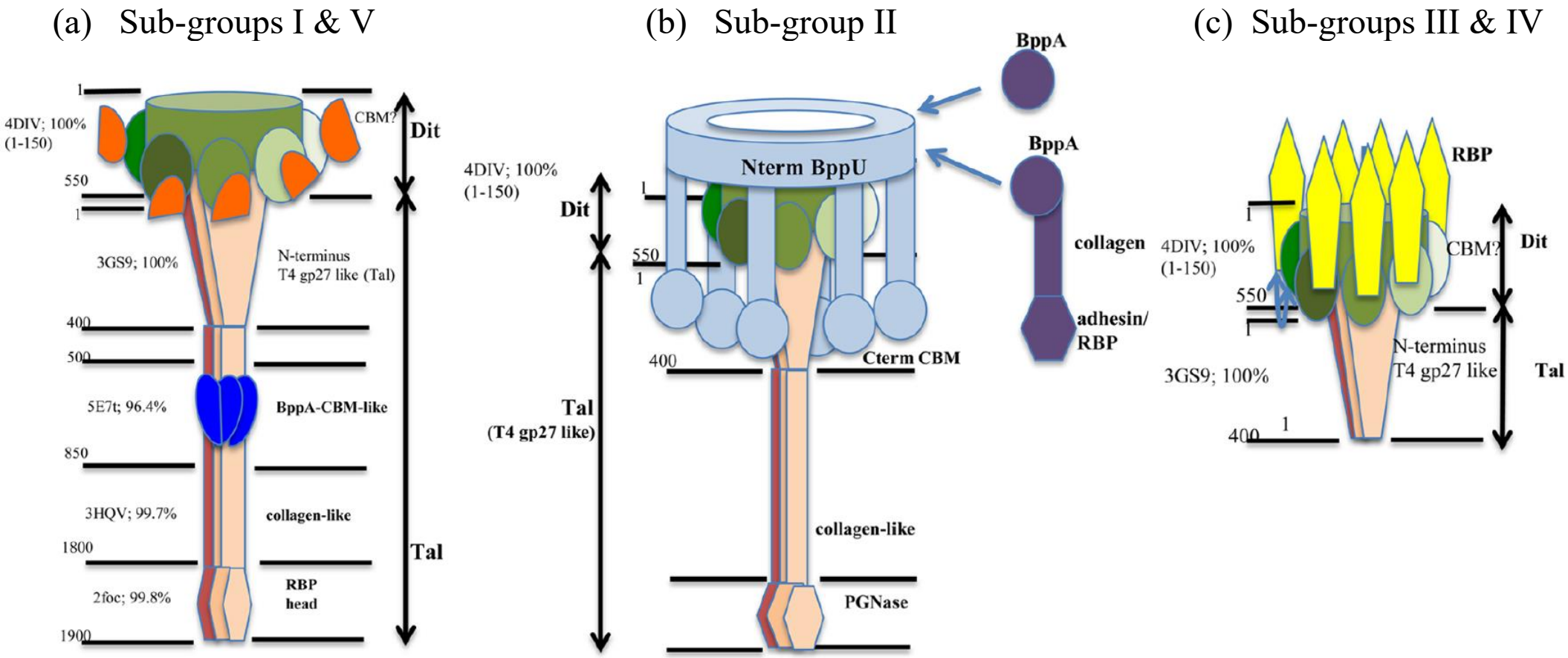


Figure 1. Schematic representation of the baseplate architecture of P335 phages belonging in (a) sub-group I & V, (b) sub-group II and (c) sub-groups III/IV. Percentages represent identity to proteins identified through HHPred analysis. Figure was adapted from Mahony *et al.* (2017) (41) with additional information from Kelleher *et al.* (2018) (53).

Table 2. Host encoded molecular receptors for bacteriophage adsorption in Gram-positive bacteria.

Host Receptor	Receptor Structures and/or Features	Bacterial Species	Bacteriophage	Bacteriophage Family	Reference
Membrane Protein	YueB	*Bacillus subtilis*	SPP1	*Siphoviridae*	(71)
	Phage infection protein (Pip)	*L. lactis*	c2	*Siphoviridae*	(58)
	YjaE	*L. lactis*	c2	*Siphoviridae*	(72)
Peptidoglycan (PG)	GlcNAc[1]/MurNAc[2] PG Residues	*Clostridium botulinum*	a2	*Siphoviridae*	(73)
	MurNAc PG Residues	*Bacillus thuringiensis*	Bam35	*Tectiviridae*	(74)
	O-acetyl groups on MurNAc	*Staphylococcus aureus*	ϕ11	*Siphoviridae*	(75)
Wall Teichoic Acid (WTA)	GlcNAc/Rha[3]-decorated poly(Rho-P)[4] WTA	*Listeria monocytogenes*	P35	*Siphoviridae*	(76)
	Rha-decorated poly(Rho-P)	*Ls. monocytogenes*	A118	*Siphoviridae*	(76)
	β-*O*-GlcNAc-decorated poly(RboP) WTA	*S. aureus*	66	*Podoviridae*	(77)
	Poly(RboP) WTA backbone	*S. aureus*	812	*Myoviridae*	(78)
Lipo-Teichoic Acid (LTA)	D-alanine esters / α-glucose residues on poly(Gro-P)[5] LTA	*Lactobacillus delbrueckii*	LL-H	*Siphoviridae*	(79)
Cell-Wall Polysaccharide (CWPS)	Hexaphosphosaccharidic CWPS	*L. lactis*	1358	*Siphoviridae*	(57)
	Hexaphosphosaccharidic CWPS	*L. lactis*	sk1	*Siphoviridae*	(55)
	Pentaphosphosaccharidic CWPS	*L. lactis*	LC3	*Siphoviridae*	(55)
	Rhamnose-rich heptasaccharide	*Lactobacillus casei*	J-1	*Siphoviridae*	(80)
Flagellum	Flagellum filament	*B. subtilis*	PBS1	*Myoviridae*	(81)
	Flagellum filament	*Bacillus pumilis*	PBP1	*Siphoviridae*	(82)

Footnote: [1]GlcNAc = *N*-acetyl-glucosamine, [2]MurNAc = *N*-acetyl-muramic acid, [3]Rha = rhamnose, [4]poly(Rho-P) = poly-ribitolphosphate, [5]poly(Gro-P) = poly-glycerolphosphate

1.3. Lactococcal cell wall glycopolymers

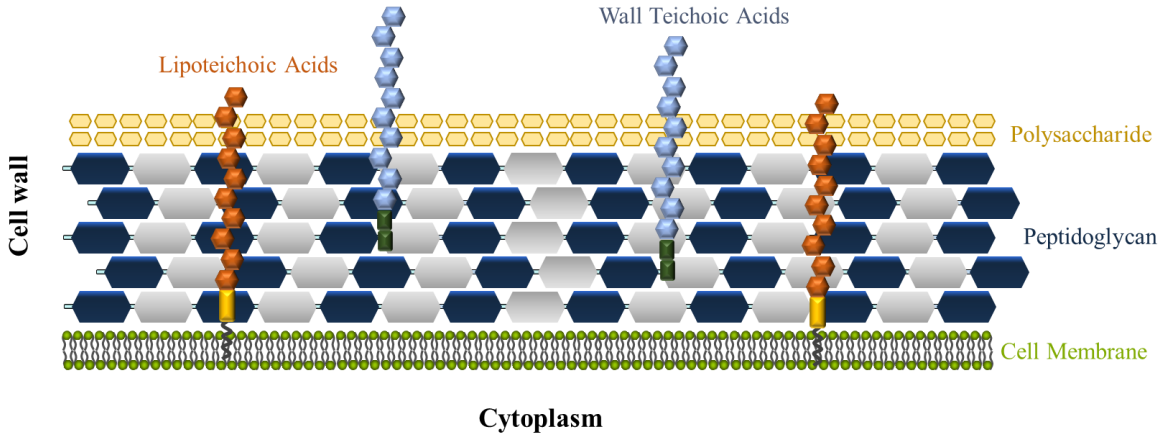


Figure 2. Schematic representation of the lactococcal cell wall illustrating major components of interest.

1.3.1 Rhamnosyl cell wall-associated polysaccharides

It is well established that polysaccharides can be involved in the process of phage adsorption onto the bacterial cell envelope of Gram negative cells. For example, coliphages T4 (83, 84) and T5 (85, 86) bind reversibly to lipopolysaccharides present on the *Escherichia. coli* cell envelope prior to irreversible adsorption to a cell membrane protein and/or other glycopolymers. Seminal work on the *B. subtilis*-infecting phage SPP1 highlighted glycopolymers as potential phage receptors in Gram-positive bacteria. SPP1 was shown to first recognise and reversibly bind to glucosylated WTAs before binding irreversibly to the integral membrane protein, YueB (71) (Figure 2). Original work on *L. lactis* phages indicated similar types of interactions to glycopolymers based on competitive inhibition assays with monosaccharides such as rhamnose, glucose, and galactose to limit the infectivity of phages such as kh (936-type), sk1 (936-type), and others (58, 87, 88). Dupont *et al*. (2004) identified bacterial host genes involved in lactococcal 936 phage adsorption through the analysis of bacteriophage insensitive mutants (BIMs) of strains *L. lactis* IL1403 and Wg2 (54). Consequently the genetic components involved in the biosynthesis and assembly of the *L. lactis* cell wall-associated polysaccharide (CWPS, Figure 2) were identified within a gene cluster that

contained, among others, genes predicted to encode glycosyltransferases and enzymes involved in sugar precursor production (54). Similar gene clusters have been identified in all currently available *L. lactis* genomes, albeit with distinct sequence and organisational differences (55, 65).

A conserved characteristic of this gene cluster appears to be the presence of *rmlABCD* homologues (55, 65). This set of genes is known to be involved in the production deoxythimidine diphosphate (dTDP)-L-rhamnose from glucose-1-phosphate and deoxythymidine triphosphate (dTTP) (89, 90). Futhermore, the *cwps* gene cluster appears to be organised into two distinct regions if the sequence identity of different *L. lactis cwps* gene cluster is taken into consideration. The 5'-end of the gene cluster, which also contains *rmlABCD*, is highly conserved amongst all *L. lactis* genomes, while the 3'-end is characterized by a much higher degree of sequence diversity and gene content disparity (55, 65). Preliminary research on the *cwps* gene cluster sequences of six fully-sequenced strains of *L. lactis* determined that three *cwps* genotypes can be distinguished based on the DNA content of the 3'-end of *cwps*, named *cwps* type A, B, and C. A multiplex PCR-based typing method was devised using the unique regions found within each *cwps* type (65). Since then the list of available *L. lactis* genomes has expanded (to more than 30), however, this genotyping scheme based on the *cwps* gene cluster can be still successfully applied. Focused analysis of *L. lactis* strains belonging to *cwps* type C identified further subdivision within this group with five subtypes established, named *cwps* type C_1 – C_5 (55). An outline of the genomic comparison between the three types of CWPS biosynthesis gene clusters can be seen in Figure 3. As mentioned above, CWPS structures, and by extension the genotypes these structures are associated with, have been shown to play a crucial role in phage/host interactions whereby certain 936 phage RBP subtypes only appear to infect strains within a single CWPS genotype (or at most two CWPS genotypes) (40, 65). In particular, when a set of genes found within the 3'-end region of the *cwps* gene cluster was mutated in *L. lactis* NZ9000 (a derivative of *L. lactis* MG1363) and complemented by their homologs/analogs found in *L. lactis* 3107, the

resulting strain was shown to produce the CWPS structure of *L. lactis* 3107, while the recombinant strain had also acquired the phage sensitivities of *L. lactis* 3107 (55).

CWPS structures have been determined for a number of *L. lactis* strains. The CWPS of *L. lactis* MG1363 was the first such structure to be defined among lactococcal strains and was shown to consist of two components. The first component to be identified and characterised was a polymer of repeating hexasaccharide subunits joined by phosphodiester bonds, a structure that was referred to as the pellicle (Table 3)(56). More recently, an additional component of the *L. lactis* MG1363 CWPS was identified which is closely associated and "buried" beneath the pellicle, composed almost exclusively of rhamnose (Rha) residues, and thus termed the "rhamnan" (Table 3)(91). The rhamnan polymer essentially consists of a trisaccharide repeat unit of Rha and is thought to be enmeshed in the PG layer of the cell. This saccharide becomes exposed at the cell surface in mutant strains of *L. lactis* lacking the polysaccharide pellicle layer (91). In another lactococcal strain, *L. lactis* 3107 (*cwps* type C_2), the rhamnan component was identical to that of *L. lactis* MG1363 (*cwps* type C_1) while the structure of the pellicle differed (55, 91). In *L. lactis* 3107, the pellicle is composed of phospho-pentasaccharide repeat units, containing glucose (Glc), N-acetylglucosamine (GlcNAc), and galactofuranose (Gal*f*) (Table 3) (55). Together the rhamnan and the pellicle form the CWPS of type C strains that is assembled through the enzymatic activities encoded by the *cwps* gene cluster mentioned above (55, 56, 65, 91). However, the nature of the presumed covalent linkage between the pellicle and the rhamnan has not yet been elucidated.

More recently, the CWPS structures of representative strains from genotypes A and B have been determined. *L. lactis* UC509.9 belongs to the A genotype (65) and exhibits a CWPS that is structurally very distinct from those of the type C strains (Table 3) (92). The CWPS of *L. lactis* UC509.9 contains a rhamnan composed of rhamnose-rich linear tetrasaccharide subunits. This neutral polysaccharide was shown to be partially substituted with a branched phosphorylated oligosaccharide. This negatively charged oligosaccharide in *L. lactis* UC509.9

is covalently attached to the rhamnan and, as such, forms a single polysaccharide structure (92). Similarly, the CWPS of the *cwps* B genotype strain *L. lactis* IL1403 was determined (93) and shown to be composed of a rhamnan backbone made from linear rhamnosyl disaccharide subunits irregularly decorated with a glycerophosphate-containing trisaccharide (Table 3) (93).

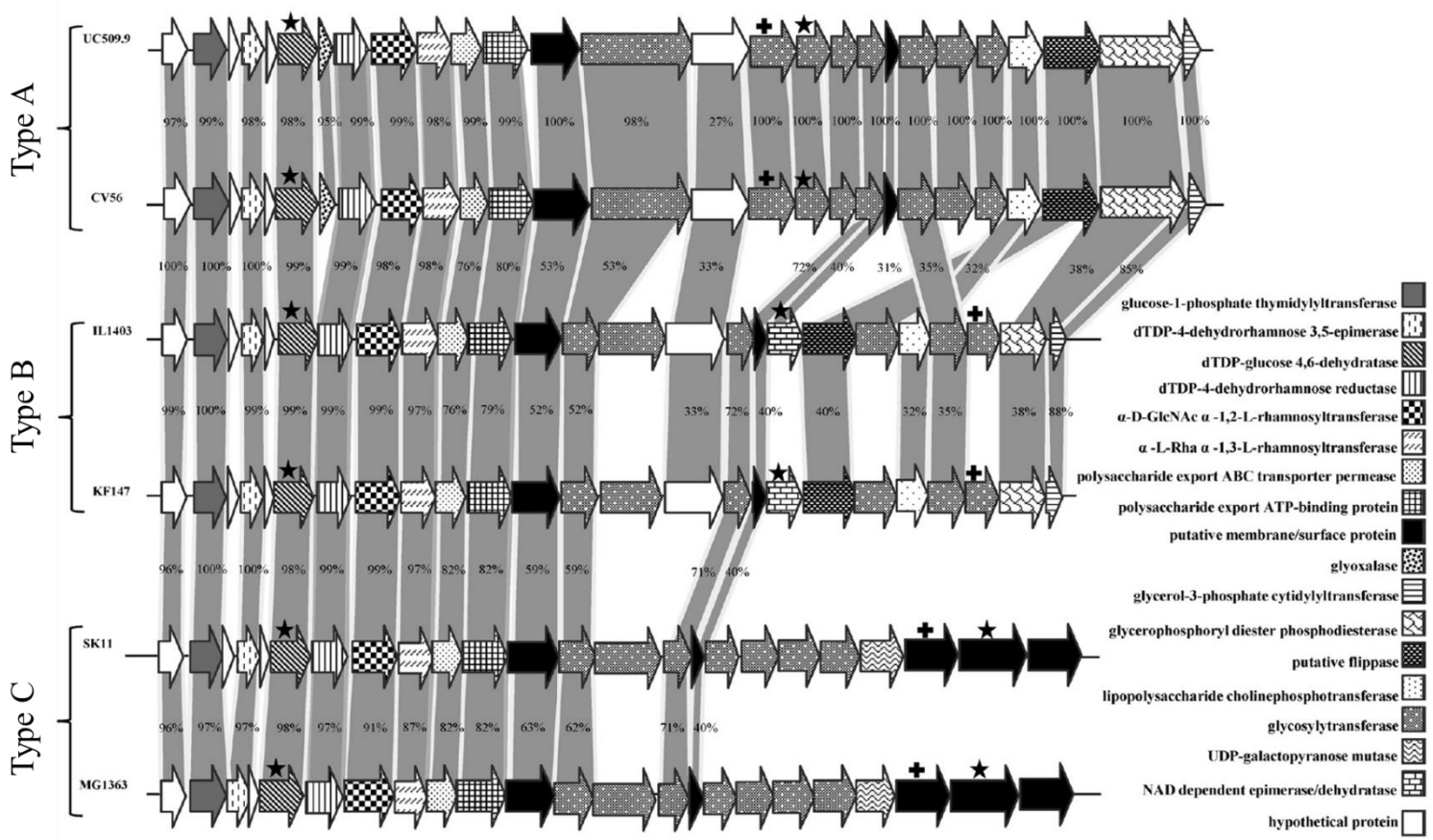


Figure 3 Comparison of the genomic regions encoding the CWPS biosynthesis cluster. This includes the genomic comparison of six *L. lactis* strains based on which the three distinct CWPS subgroups were identified. These strains included *L. lactis* UC509.9, CV56, IL1403, KF147, SK11, and MG1363. The genes highlighted by a cross/star in the 3'-end region of the cluster were used to create primers for a multiplex PCR typing method of the CWPS types. This figure originally appeared in Mahony *et al*. (2013) (65).

Table 3. Outline of known structures of cell wall-associated polysaccharides isolated from *L. lactis* strains belonging to CWPS genotypes A, B, and C.

Lactococcal Strain	CWPS Genotype	Cell wall-associated Polysaccharide Structure	Reference
UC509.9	A	-[-6-α-Glc-2-α-Rha-2-α-Rha-2-α-Rha-]-[-6-α-Glc-2-α-Rha-2-α-Rha-2-α-Rha-]- |4 β-Gal[1](6*P*-1-Gro[2])-2-α-Rha-2-α-Rha-3-β-GlcNAc |3 |6 α-Rha α-Glc	(92)
IL1403	B	-[-3-α-Rha-2-α-Rha-]-[-3-α-Rha-2-α-Rha-]- |4 β-Glc-4-α-[3]GalNAc(6*P*-1-Gro)-3-β-GlcNAc	(93)
		Rhamnan	Pellicle	
MG1363	C_1	[-2-α-*L*-Rha-2-α-*L*-Rha-3-α-*L*-Rha-]	-[-6-α-GlcNAc-3-α-Rha-3-β-GlcNAc-2-β-Galf-6-α-Glc-*P*-]- |6 α-Glc	(56, 91)
3107	C_2	[-2-α-*L*-Rha-2-α-*L*-Rha-3-α-*L*-Rha-]	-[-6-α-Glc-3-α-Galf-3-β-GlcNAc-2-β-Galf-6-α-GlcNAc-*P*-]-	(55, 91)
SMQ-388	$C_?$	*N.D.*	-[-6-α-GlcNAc-3-α-Galf-3-β-GlcNAc-2-β-Galf-6-α-GlcNAc-*P*-]- |6 α-Glc	(94)

Footnote: [1]Gal = galactose, [2]6*P*-1-Gro = glycerophosphate, [3]GalNAc = *N*-acetyl-galactosamine.

1.3.2 Biosynthesis and assembly of the CWPS

Currently, knowledge on the biosynthetic steps involved in CWPS manufacture in *L. lactis* is quite limited. Therefore, increased scientific attention is to be focused on defining and characterising the processes involved in the synthesis of this polysaccharidic structure. Interestingly, this research has provided compelling evidence to suggest that genes beyond those found within the *cwps* gene cluster and their products also contribute to the final CWPS structure. The initiating step of rhamnan precursor biosynthesis involves the priming of an undecaprenyl-phosphate lipid carrier in the bacterial membrane with a sugar residue, commonly a GlcNAc-1-P (91, 95), similar to the initiation of WTA biosynthesis in both *B. subtilis* and *S. aureus* (96). Such enzymes have been identified and experimentally proven to be involved in CWPS biosynthesis in several Gram-positive bacteria, such as RgpG in *Streptococcus mutans* (97) and GbsO in *Streptococcus agalactiae* (98). A protein with high sequence similarity to these enzymes was identified in *L. lactis* MG1363, named TagO, and is believed to be involved in the assembly of the lactococcal rhamnan as conditional mutants of *tagO* in *L. lactis* MG1363 exhibit severely attenuated growth, reduced amounts of Rha in their cell wall and a distinctly altered cell morphology (91). The final attachment of the completed CWPS onto the PG layer has also been investigated in *L. lactis*. In other Gram-positive microorganisms, including those belonging to the genera *Bacillus* and *Staphylococcus*, enzymes of the LytR-CpsA-Psr (LCP) family are believed to covalently attach glycopolymers onto the PG layer (99-101). In the case of *L. lactis* NZ9000, it has been suggested that an *lcp* paralog, *lcpA*, is involved in the covalent attachment of the completed rhamnan molecule onto the PG since deletion of this gene results in the rhamnan component being more loosely associated with the PG (91). Interestingly, neither *tagO* nor *lcpA* are located within the *cwps* gene cluster (91).

As mentioned above, enzymes encoded by genes in the *cwps* gene cluster play a major role in the biosynthetic processes of both the rhamnan component as well as the anionic

polysaccharide decorations of the rhamnan. However, functional assignments and accurate annotations of the genes within this gene cluster have not been undertaken to date. Despite this lack of experimental evidence in relation to the genes in the lactococcal *cwps* gene cluster, research on the biosynthetic assembly of streptococcal CWPS has provided a compelling model for the process of rhamnan assembly, which can be divided in four to five steps (Figure 4). These steps include: (1) **initiation** of polysaccharide assembly through the function of enzymes such as TagO, (2) **elongation** of the polysaccharide subunit or precursor molecule through the step-wise addition of sugar residues catalysed by dedicated glycosyltransferases, (3) **translocation** of the completed subunit or precursor molecule across the cell membrane through the action of a dedicated ABC-type (ATP binding cassette) transporter or flippase, (4) **linkage** of the polysaccharide molecule onto the PG through the action of Lcp-type enzymes, (5) and finally possible **secondary modifications** that may occur after the attachment of the (polymerised) molecule onto the PG layer (95). Focused analysis of the gene products of streptococcal *cwps* gene clusters has identified three glycosyltransferases, RgpA, RgpB, and RgpF, that are involved in the production of the streptococcal CWPS and more specifically in the assembly of the rhamnosyl backbone of the structure (102). Interestingly, genes with sequence identity to those of RgpABF were identified at the 5'-end of the lactococcal *cwps* gene cluster, which provides further evidence for the proposed genetic organisation of the cluster in terms of rhamnan/pellicle biosynthesis. Even though within the lactococcal *cwps* gene cluster there are gene products whose predicted function could theoretically assemble the complete CWPS structure, the specificity and function of these enzymes still needs to be experimentally validated. Finally, as will be discussed in more detail in Chapters 3 and 4 of this thesis, in addition to the above described process for rhamnan assembly, we have identified and described two additional modes of CWPS assembly/modification. These include the process by which the pellicle is assembled and attached onto the rhamnan, as well as the process for extra-cytoplasmic glycosylation of the rhamnan or pellicle.

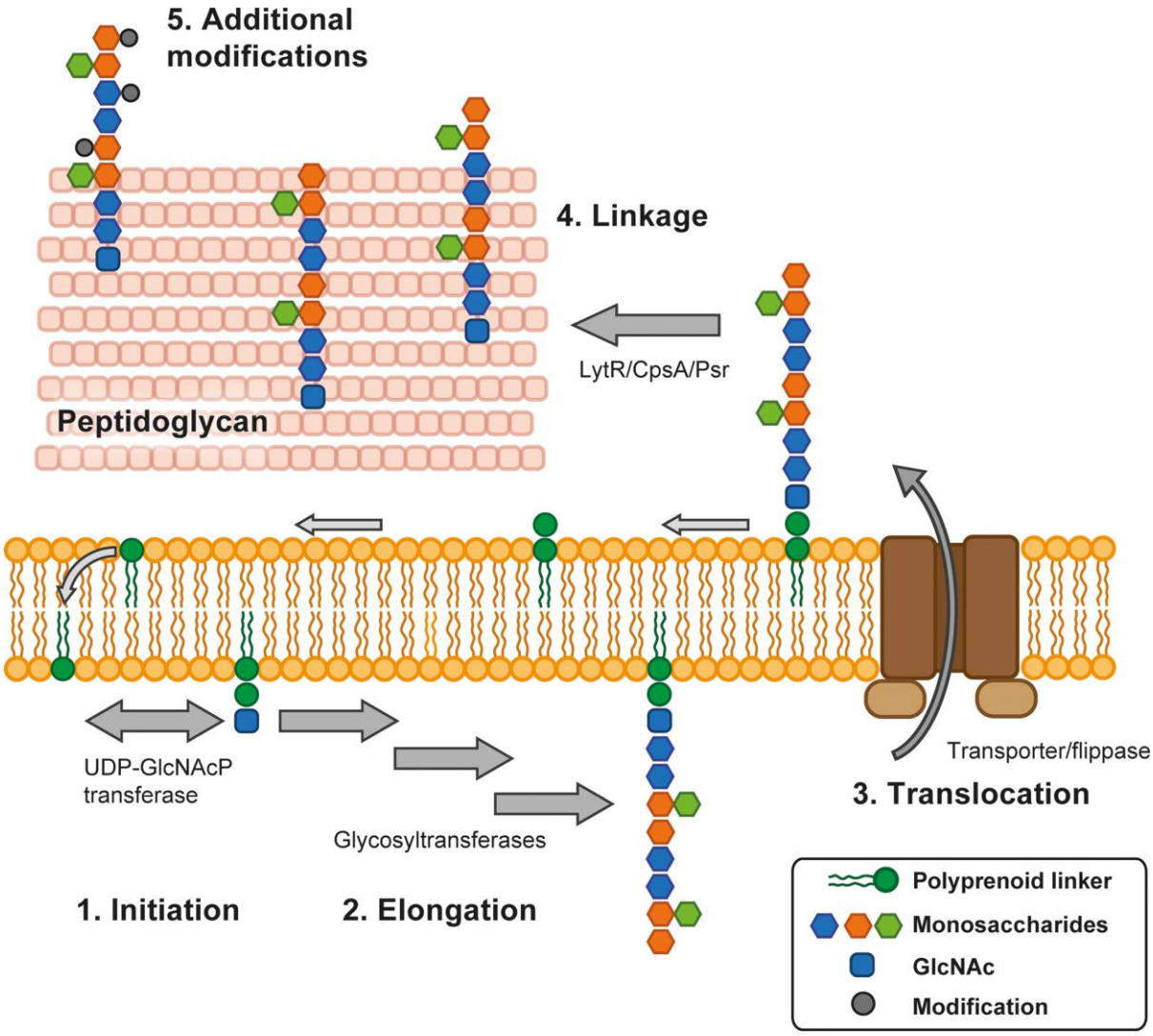


Figure 4 Proposed steps involved in the biosynthesis of streptococcal glycoconjugates, such as the RhaCWPS. Figure first appeared in Mistou *et al*. (2016) (95)

1.4. Lactococcal teichoic acids

1.4.1 Wall-teichoic acid structure and assembly

In addition to the CWPS that is associated to the lactococcal PG layer, another anionic glycopolymer can be found in the cell envelope of Gram-positive microorganisms, known as teichoic acid (TA) (96, 103). Two main varieties of this molecule are known. One is attached to a lipid and therefore anchored in the plasma membrane of the Gram-positive cell, commonly referred to as the lipo-teichoic acid (LTA, more on this structure in the following section). The other is covalently attached to the PG layer and is known as the wall teichoic acid (WTA) (96, 103). LTA and WTA represent anionic macromolecules due to the large number of incorporated phosphates in their structure (96) (Figure 2); however, thus far *L. lactis* has only been shown to produce LTA (104, 105).

Our knowledge on the WTA structure and its biosynthetic pathway is largely based on research of model bacterial species such as *B. subtilis* and *S. aureus*, although WTA has been identified in several LAB (96, 103). As mentioned above WTA is associated with the PG layer through a phosphodiester linkage to the *N*-acetyl-muramic acid (MurNAc) of the glycan chain (96). The structure of a WTA can be divided into two main parts. The first, known as the "linkage unit", is directly connected to the PG and is most commonly composed of a *N*-acetyl-mannosamine (ManNAc) β1→4 GlcNAc disaccharide followed by a set of one to three glycerol/ribitol phosphate residues (see below) attached to the ManNAc side of the disaccharide (Figure 5) (96). The composition of this linkage unit is highly conserved among bacterial species (103). The WTA linkage unit is followed by a long chain of either glycerol 3-phosphate (GroP) or ribitol 5-phosphate (RboP) residues, known as the "main chain", that is the source of the macromolecule's anionic charge (Figure 5) (96). Nonetheless, WTAs that incorporate more unique polyol-based residues have also been identified (103, 106, 107). The incorporation of a poly-GroP, a poly-RboP, or a combination or variation of those residues in the main chain is strain-specific as is the specific length of the polyol chain (96, 103, 108-110).

Finally, the main chain of WTA is decorated in certain cases with D-alanine esters (D-ala) and/or monosaccharides such Glc, GlcNAc, and galactose (Gal) (96).

Since both poly-GroP and poly-RboP WTAs attach to the same type of linkage unit, the initial biosynthetic steps of their construction are identical while they diverge in subsequent stages (Figure 6) (96). The genes that encode the proteins responsible for the production of poly-GroP WTAs are known as *tag* genes (teichoic acid glycerol) while those required for the production of poly-RboP WTAs are known as *tar* genes (teichoic acid ribitol) (96) and both were described in *B. subtilis*. The first step involves the transfer of an activated GlcNAc residue onto a membrane associated undecaprenyl phosphate (UndP) by TagO/TarO, a process that mirrors the assembly of Gram-positive CWPS (95). TagA/TarA transfers a ManNAc residue onto the GlcNAc-UndP forming the disaccharide molecule of the WTA linkage unit. Finally, the ManNAc residue is primed with a single glycerol-phosphate unit through the action of TagB/TarB (96, 111, 112). These initial three steps are identical in the biosynthetic pathways of both poly-GroP and poly-RboP WTAs. The production of poly-GroP WTA continues with the processive attachment of glycerol-phosphate units through the polymerising action of the enzyme TagF (113, 114). TagB and TagF therefore work together as primase/polymerase pair, respectively, both required for the poly-GroP WTA assembly. The two-component ABC transporter TagGH then transfers the fully assembled poly-GroP polymer to the extracellular side of the cell membrane (115). The final attachment of the polymer onto the PG is thought to be catalysed by three redundant enzymes, TagTUV, that belong to the LCP family (see section 1.3.2 Biosynthesis and assembly of the CWPS) (101). TarF catalyses the attachment of a single glycerol-phosphate in poly-RboP WTA production in *B. subtilis* (96, 103). The primase/polymerase function performed by the TagB/TagF pair is performed in poly-RboP WTA biosynthesis by TarK and TarL, respectively (96, 116, 117). Following polymerisation of the poly-RboP WTA, the same biosynthetic steps as those involved in the poly-GroP WTA production are taken (96, 103). A representation of these steps is presented in Figure 6.

Interestingly, while *S. aureus* exhibits a poly-RboP WTA as is the case for *B. subtilis*, the two species employ a distinct biosynthetic pathway for its assembly (Figure 6) (118, 119). As mentioned above, the priming and polymerisation of the main chain in *B. subtilis* is performed by the enzymatic pair TarK/TarL, however in *S. aureus*, the same process is performed by two distinct enzymes independently, also named TarK and TarL (116-118). Thus in *S. aureus*, TarK and TarL both function as primases and polymerases independently and concurrently, each producing a distinct poly-RboP WTA polymer of slightly different length and structure (103). The remaining steps involved in the assembly of the *S. aureus* poly-RboP WTA remain the same as in *B. subtilis*.

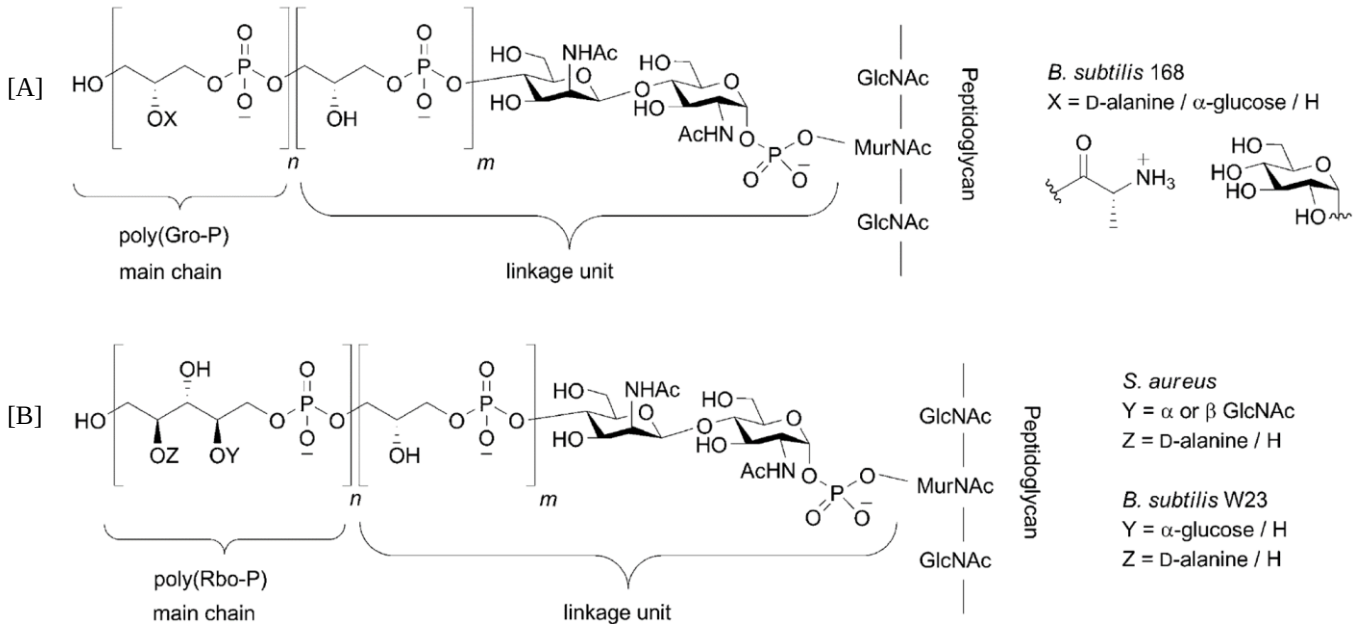


Figure 5. Representative structures of wall-teichoic acids with [A] a poly-GroP main chain or [B] a poly-RboP main chain. Figure first appeared in Swoboda *et al.* (2010) (96)

[A]

[B]

[C]

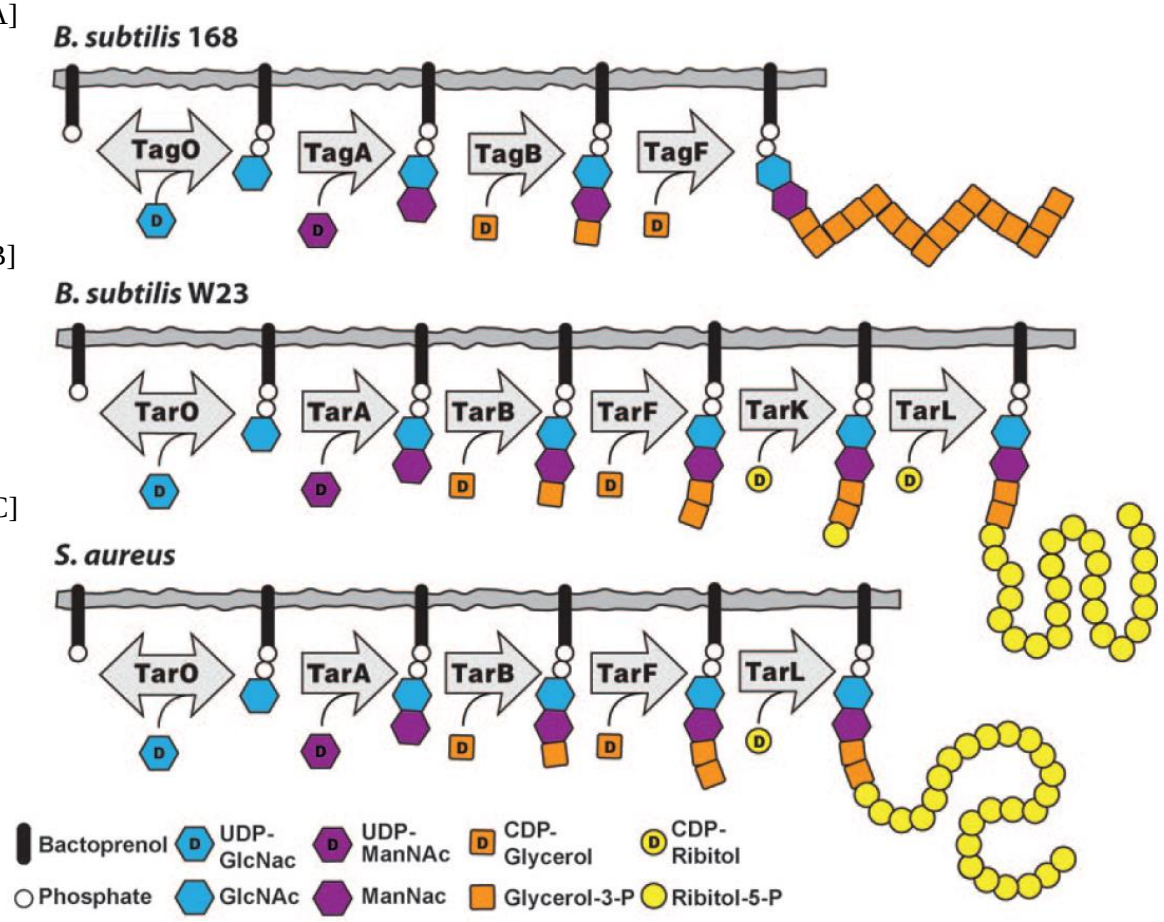


Figure 6. Schematic representation of the biosynthetic steps involved in the WTA assembly of three Gram-positive bacteria along with the enzymes involved in each. [A] *B. subtilis* 168 [B] *B. subtilis* W23 [C] *S. aureus*. Figure first appeared in Swoboda JG, *et al.* (2010) (96)

1.4.2 Lipo-teichoic acid structure and assembly

Like WTA, lipo-teichoic acid is anionic glycopolymer that typically consists of GroP repeating units (Figure 7) (120, 121) that may be decorated with D-Ala, Glc, GlcNAc or Gal residues although D-Ala is the most commonly encountered GroP substitution (120). In LTAs, this poly-GroP chain is connected to the lipid membrane via a glycolipid anchor. The glycolipid anchor is commonly a disaccharide glycolipid, such as a pair of Glc and/or Gal residues connected to diacylglycerol (DAG) (120, 121). Interestingly, a wide variety of LTAs have been identified and characterised in Gram-positive bacteria and these glycopolymers can be modified between species in several ways. For example, different bacterial species, and even strains, have been shown to vary the degree and type of poly-GroP chain substitutions (122, 123). Furthermore, unique main chain anionic repeat units have been identified, with some even incorporating sugar residues in each repeat unit (124-126). LTAs can vary in the type of glycolipid anchor they use. For example, some LTAs can vary in the number of sugar residues attached to the DAG as well as incorporating substitutions onto these sugar residues (127-129). The LTA structures found in *L. lactis* exhibit a Glc_2-DAG (and in some cases Glc-AcylGlc-DAG) glycolipid anchor, a poly-GroP main chain decorated with mostly D-Ala and Gal substitutions (105, 130-132).

The first step in LTA biosynthesis is performed by the glycosyltransferase YpfP through the addition of two Glc residues onto the DAG cell membrane molecule (133). Once assembled, the glycolipid anchor is "flipped" to the exterior of the cell membrane by LtaA, a polytopic transmembrane protein (134, 135). Finally, a LTA synthase with an extracellular active site, known as LtaS, processively attaches a GroP residues onto the glycolipid anchor and completes the anionic polymer (136).

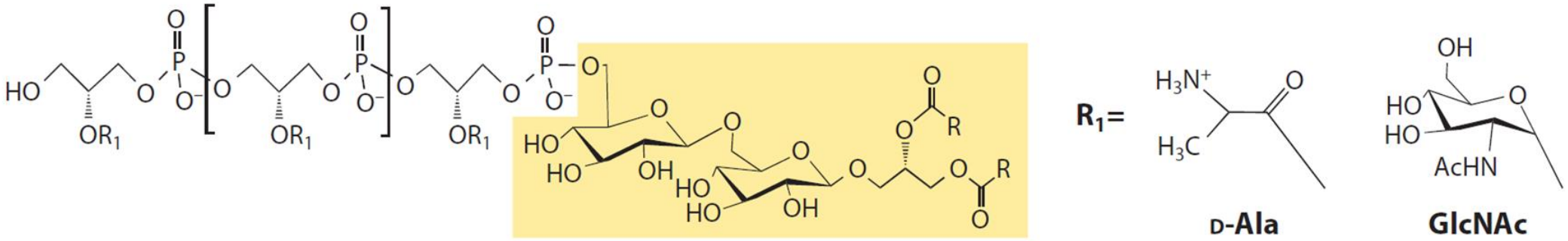


Figure 7. Schematic representation of a common Gram-positive LTA structure. Yellow-shaded region = Glycolipid Anchor, R=Fatty Acids. Figure originally appeared in Percy and Gründling (2014) (121).

1.4.3 Modification of teichoic acids

One of the most important aspects of teichoic acid structure and functionality is intrinsically related to the molecular substitutions found attached to the glycopolymer's main chain. The most prominent and widely studied type of main chain modification is the addition of D-Ala esters onto the extending poly-GroP (103, 120, 121). The degree of teichoic substitution with D-Ala may affect the TA's overall charge and hydrophobicity, due to the positive charge of these residues. The main steps involved in D-alanylation take place in the extracellular part of the cell membrane and are thought to be the same for both WTA as well as LTA production (103, 120, 121). The genes involved in this process are known as *dltABCD* and have been identified in the genomes of a large number of Gram-positive bacteria (137). The main biosynthetic steps that are catalysed by the *dlt*-encoded enzymes are highlighted in Figure 8. Briefly, DltA is a cytoplasmic protein thought to activate a D-Ala using an ATP molecule and further coupling this activated residue to DltC. Although the function of both DltB and DltD has not been completely elucidated, DltB is thought to transfer the D-Ala from DltC onto an UndP molecule in the cell membrane and from there transfer the coupled D-Ala-P-UndP to the extracellular region. Finally, DltD is thought to catalyse the final attachment of D-Ala onto TA (103, 120, 121).

The degree of D-alanylation is variable and a number of different environmental conditions have been implicated in controlling the extent of this process in Gram-positive bacteria. For example, increased pH (138), sodium chloride (NaCl) concentration (139), and temperature (140) all lead to a decrease of D-alanylation. Furthermore, mutant strains lacking D-Ala residues on their TA exhibit increased susceptibility to host defence mechanisms such as phagocytosis (141). Finally, increased D-alanylation of *L. lactis* and Group B *Streptococci* LTA has been linked to increased resistance to cationic antimicrobial compounds, such as nisin,

presumably due to the finding that D-Ala substitutions provide increased cell wall density (142, 143). In general, D-alanylation appears to exert a significant effect on the overall phenotypic nature and environmental response of lactococcal species.

Despite the extensive catalogue of both WTA and LTA structures from a diverse array of Gram-positive species, as well as our knowledge for the bio-machinery involved in their assembly, rather little is known about the importance of glycosylation of TAs and the genetic and molecular mechanisms underpinning it. However, major steps have been taken to highlight and understand these processes (121, 144-148). Unlike TA alanylation in which similar pathways are involved in WTA and LTA substitution, glycosylation is suggested to follow a distinct biosynthetic pathway in the two TA subcategories.

The process of LTA glycosylation closely resembles a commonly encountered three-component system known to be involved in the glycosylation of various glycomolecules in bacteria (149). This system utilises two independent glycosyltransferases and a dedicated flippase enzyme. In the first step of the process an intracellularly-acting, cell membrane-associated glycosyltransferase couples an activated sugar residue to an UndP molecule. This is followed by the re-orientation of the sugar-activated UndP to the extracellular side of the cell membrane through the activity of a flippase, while an integral membrane-associated glycosyltransferase then transfers the sugar from the UndP and attaches it to the LTA (121). A group of enzymes that serve such functions have recently been identified in *S. aureus*, namely CsbB (UndP sugar-activating enzyme), GtcA (putative flippase), and YfhO (LTA glycosyltransferase), and were shown to GlcNAcylate this strain's LTA. Furthermore, the genes encoding these enzymes were shown to be overexpressed in stress-inducing conditions, such as increased concentration of salt in the growth media (150). It has furthermore been reported that a galactosylated LTA in *L. lactis* confers resistance to bacteriophage infection as opposed to a non-galactosylated version of the same molecule (105).

The glycosylation of WTA has been thought to occur exclusively on the intracellular side of the cell membrane, once the whole polymer has been assembled and prior to its extracellular transport. This process is thought to be accomplished through the activity of single glycosyltransferases. Enzymes that have been identified as WTA glycosyltransferases include TagE, in *B. subtilis* 168, that decorate this strain's WTA with glucose residues (146), TarQ that is involved in the glycosylation of *B. subtilis* W23 poly-RboP WTA (147), and enzymes such as TarM and TarS that attach GlcNAc residues onto *S. aureus* WTA (77). However, recently, a three-component glycosylation system similar to the one involved in LTA glycosylation has been implicated in the sugar substitution of *Ls. monocytogenes* WTA (148), indicating that this might be a multi-substrate glycosylation system not limited to a specific type of TA. Furthermore, the extent of WTA glycosylation appears to impact on phage infectivity (77), resistance to β-lactam antibiotics, such as methicillin (147), and virulence (148).

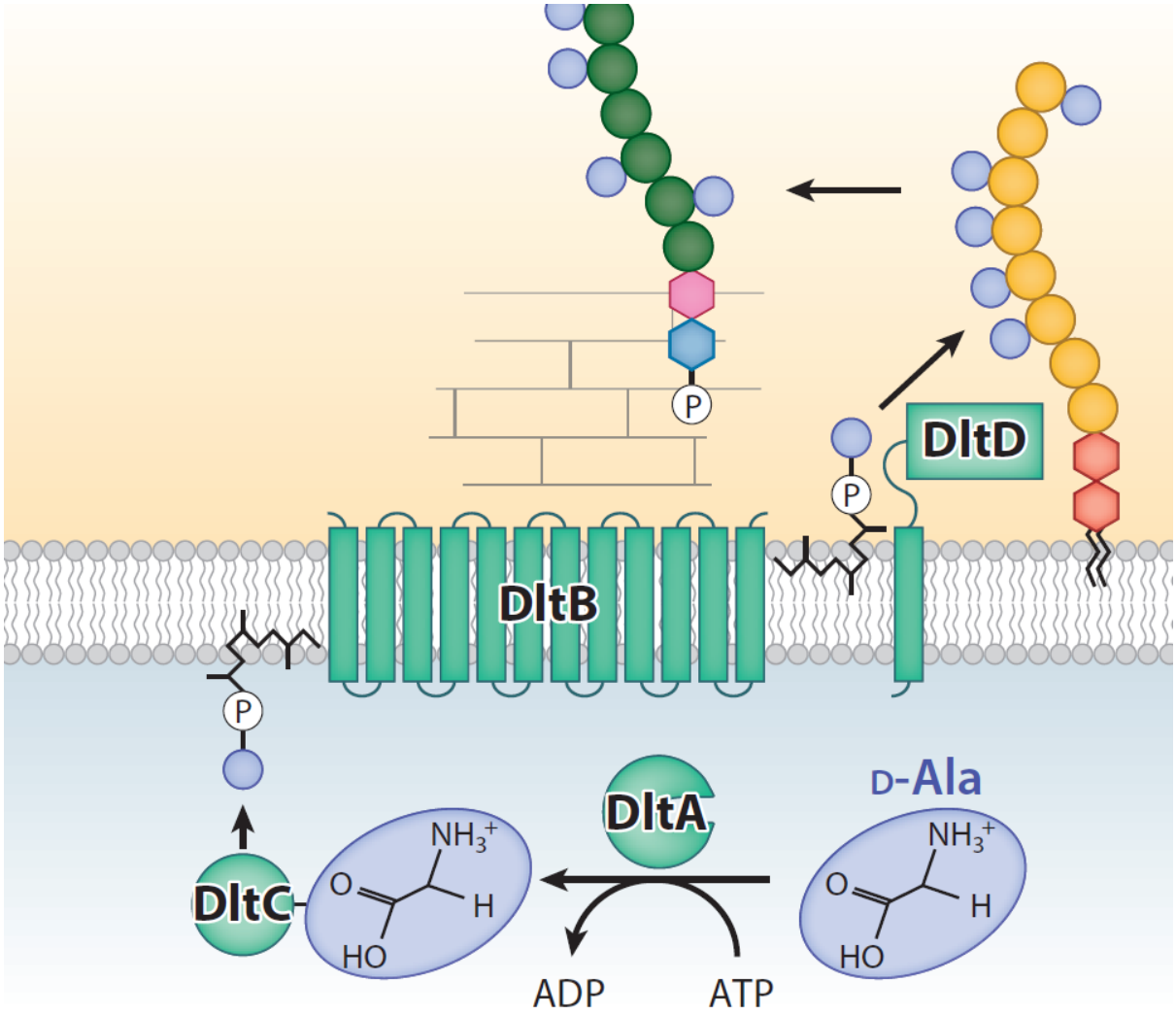


Figure 8. Schematic representation of the molecular steps involved in the D-alanylation of TA, in this case LTA specifically, as well as the enzymes that perform them. Figure originally appeared in Percy and Gründling (2014). (121)

1.5. Genomic modification of lactoccocal strains

Our understanding of the molecular pathways involved in the assembly of all of these glycopolymers has relied heavily on our ability to genomically modify LAB and specifically lactococcal strains. Over the years, a number highly successful, although relatively cumbersome techniques have been developed and used for the purpose of mutational analysis of *L. lactis* (151-153). The majority of them follow a strategy of homologous double-crossover recombination, that relies on RecA-dependent crossover event and the use of plasmids, which may be nonreplicative or conditionally replicative. Counterselection markers have been implemented in several varitions of this technique, however, mutant isolation remains quite laborious (154). Another more recent technique, known as recombineering, relies on prophage-derived recombinases, such as Redβ and RecT, to incorporate single-stranded DNA (ssDNA) into the genome of the strains of interest (155, 156). A small number of studies have successfully implemented the latter technique to study CWPS morphogenesis (55). However, a major drawback of such a technique includes to the absence of any selectable marker and its heavy reliance of high through-put colony PCR screening.

In the last decade, the discovery of the clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 system has led to an upsurge of novel mutagenesis techniques that take advantage of this system's precision. Recently, ssDNA recombineering has been succesfully combined with the CRISPR/Cas9 in order to introduce desired mutations in *Lactobacillus reuteri* (157, 158) (Figure 9). The recombineering system as previously described (156), remains unchanged in this technique. However, the CRISPR/Cas9 system serves as counterselectable marker, whereby the Cas9 nuclease is programmed to incorporate double-stranded breaks and thus eliminate the wild-type subpopulation allowing for easier selection of the strains with desired mutation. In this study, we have adapted this CRISPR/Cas9-assisted recombineering for use with *L. lactis*.

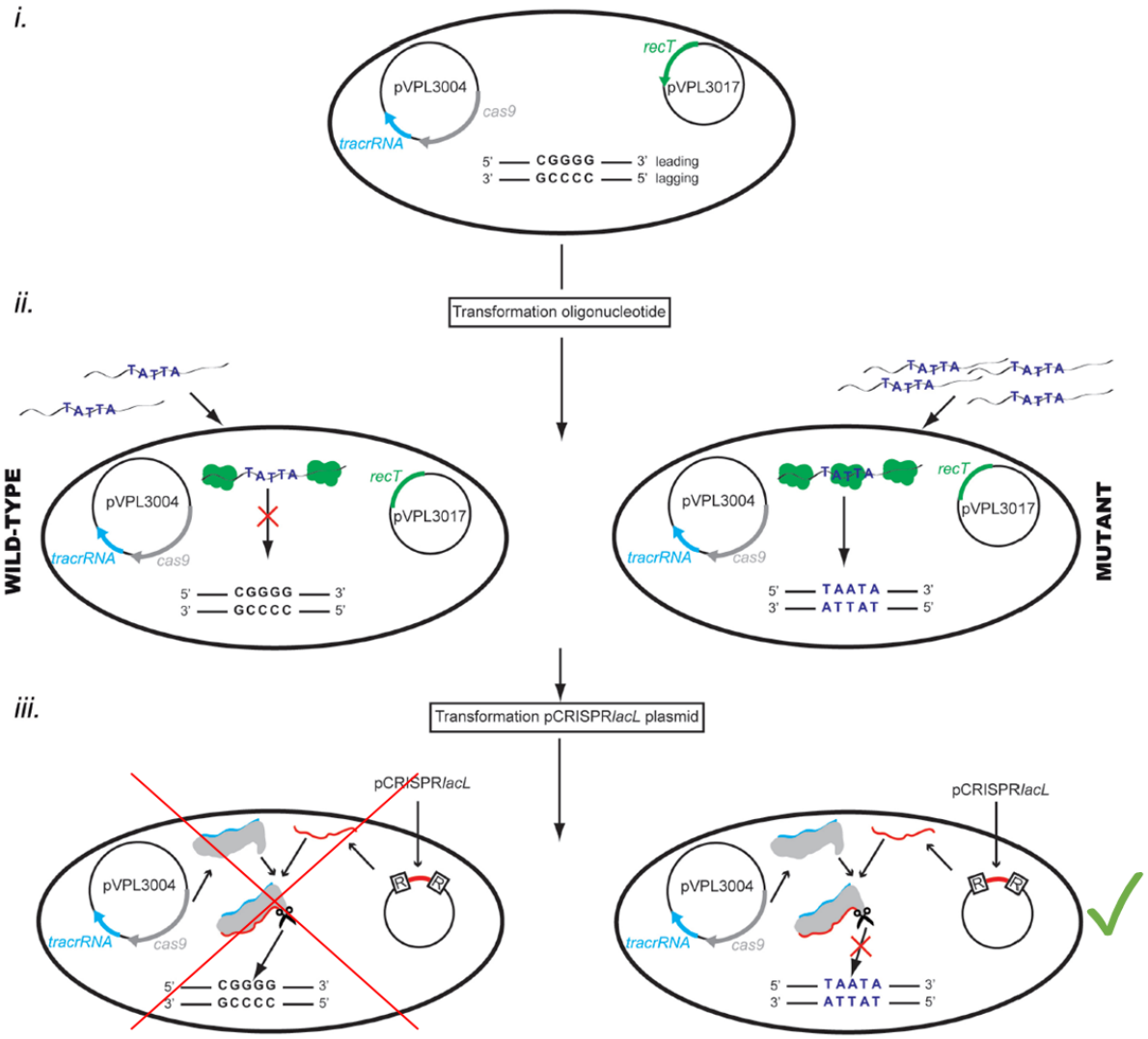


Figure 9. Schematic representation of the CRISPR/Cas9-asssisted recombineering previously developed for targeted genomic modification of *Lactobacillus reuteri*. Figure originally appeared in Oh and van Pijkeren (2014) (157).

1.6. Conclusions and outline of this thesis

The importance of lactococcal cell wall-associated glycopolymers, such as the CWPS, as major receptors for phage adsorption has been extensively established in recent literature. This thesis is aimed at expanding our current understanding of the genetic and molecular machinery responsible for the assembly of the glycopolymers. Firstly Chapter 2 will focus on the determination of the genes responsible for the attachment of glucose decorations on the pellicle of strains such as *L. lactis* NZ9000 and MG1363. The transcriptional pattern of these genes is also investigated along with the specificity of the encoded enzymes. Next, in Chapter 3, through extensive investigation of putative glycosyltransferase systems encoded by a large selection of lactococcal genomes, multiple versions of a previously described three-component glycosyltransferase system will be identified and characterised. In Chapter 4 the specificities and functionalities of a subset of the *cwps* genes will be investigated through the directed knock-out mutagenesis of each gene by employing an approach that combines recombineering as well as CRISPR/Cas9 techniques, and the characterisation of the modified CWPS structures of the *cwps* gene mutants. Finally Chapter 5 discusses the isolation and characterisation of phage escape mutants, being derived from a lytic 936-type phage, that are able to re-infect previously resistant lactococcal strains exhibiting a pellicle-negative phenotype.

1.7. References

1. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welker D, Hughes J, Goh Y, Benson A, Baldwin K, Lee JH, Díaz-Muñiz I, Dosti B, Smeianov V, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Parker C, Breidt F, Broadbent J, Hutkins R, O'Sullivan D, Steele J, Unlu G, Saier M, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D. 2006. Comparative genomics of the lactic acid bacteria. PNAS 103:15611-15616.
2. Ross P, Morgan S, Hill C. 2002. Preservation and fermentation: past, present and future. International Journal of Food Microbiology 79:3-16.
3. Douillard FP, de Vos WM. 2014. Functional genomics of lactic acid bacteria: from food to health. Microbial Cell Factories 13 Suppl 1:S8.
4. Settanni L, Moschetti G. 2010. Non-starter lactic acid bacteria used to improve cheese quality and provide health benefits. Food Microbiology 27:691-697.
5. de Vos WM. 2011. Systems solutions by lactic acid bacteria: from paradigms to practice. Microbial Cell Factories 10:S2-S2.
6. Gaspar P, Carvalho AL, Vinga S, Santos H, Neves AR. 2013. From physiology to systems metabolic engineering for the production of biochemicals by lactic acid bacteria. Biotechnology Advances 31:764-88.
7. Mazzoli R, Bosco F, Mizrahi I, Bayer EA, Pessione E. 2014. Towards lactic acid bacteria-based biorefineries. Biotechnology Advances 32:1216-1236.
8. Wu C, Huang J, Zhou R. 2017. Genomics of lactic acid bacteria: Current status and potential applications. Critical Reviews in Microbiology 43:393-404.

9. Bongers RS, Hoefnagel MHN, Kleerebezem M. 2005. High-level acetaldehyde production in *Lactococcus lactis* by metabolic engineering. Applied and Environmental Microbiology 71:1109-1113.

10. Ladero V, Ramos A, Wiersma A, Goffin P, Schanck A, Kleerebezem M, Hugenholtz J, Smid EJ, Hols P. 2007. High-level production of the low-calorie sugar sorbitol by *Lactobacillus plantarum* through metabolic engineering. Applied and Environmental Microbiology 73:1864-1872.

11. De Boeck R, Sarmiento-Rubiano LA, Nadal I, Monedero V, Perez-Martinez G, Yebra MJ. 2010. Sorbitol production from lactose by engineered *Lactobacillus casei* deficient in sorbitol transport system and mannitol-1-phosphate dehydrogenase. Applied Microbiology and Biotechnology 85:1915-22.

12. Sybesma W, Starrenburg M, Kleerebezem M, Mierau I, de Vos WM, Hugenholtz J. 2003. Increased production of folate by metabolic engineering of *Lactococcus lactis*. Applied and Environmental Microbiology 69:3069-76.

13. Wegkamp A, van Oorschot W, de Vos WM, Smid EJ. 2007. Characterization of the role of para-aminobenzoic acid biosynthesis in folate production by *Lactococcus lactis*. Applied and Environmental Microbiology 73:2673-2681.

14. Bron PA, Kleerebezem M. 2018. Lactic acid bacteria for delivery of endogenous or engineered therapeutic molecules. Frontiers in Microbiology 9:1821.

15. Wang M, Gao Z, Zhang Y, Pan L. 2016. Lactic acid bacteria as mucosal delivery vehicles: a realistic therapeutic option. Applied Microbiology and Biotechnology 100:5691-5701.

16. Price CE, Zeyniyev A, Kuipers OP, Kok J. 2012. From meadows to milk to mucosa - adaptation of *Streptococcus* and *Lactococcus* species to their nutritional environments. FEMS Microbiology Reviews 36:949-71.

17. Kelly WJ, Ward LJH, Leahy SC. 2010. Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. Genome Biology and Evolution 2:729-744.

18. Song AAL, In LLA, Lim SHE, Rahim RA. 2017. A review on *Lactococcus lactis*: from food to factory. Microbial Cell Factories 16:15.

19. Perez T, Balcazar JL, Peix A, Valverde A, Velazquez E, de Blas I, Ruiz-Zarzuela I. 2011. *Lactococcus lactis* subsp. *tructae* subsp. nov. isolated from the intestinal mucus of brown trout (*Salmo trutta*) and rainbow trout (*Oncorhynchus mykiss*). International Journal of Systematic and Evolutionary Microbiology 61:1894-8.

20. Latorre-Guzman BA, Kado CI, Kunkee RE. 1977. *Lactobacillus hordniae*, a new species from the leafhopper (*Hordnia circellata*). International Journal of Systematic and Evolutionary Microbiology 27:362-370.

21. Rademaker JL, Herbet H, Starrenburg MJ, Naser SM, Gevers D, Kelly WJ, Hugenholtz J, Swings J, van Hylckama Vlieg JE. 2007. Diversity analysis of dairy and nondairy *Lactococcus lactis* isolates, using a novel multilocus sequence analysis scheme and (GTG)5-PCR fingerprinting. Applied and Environmental Microbiology 73:7128-37.

22. Passerini D, Beltramo C, Coddeville M, Quentin Y, Ritzenthaler P, Daveran-Mingot M-L, Le Bourgeois P. 2010. Genes but not genomes reveal bacterial domestication of *Lactococcus Lactis*. PLOS ONE 5:e15306.

23. Kelly W, Ward L. 2002. Genotypic vs. phenotypic biodiversity in *Lactococcus lactis*. Microbiology 148:3332-3.

24. Wegmann U, O'Connell-Motherway M, Zomer A, Buist G, Shearman C, Canchaya C, Ventura M, Goesmann A, Gasson MJ, Kuipers OP, van Sinderen D, Kok J. 2007. Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. Journal of Bacteriology 189:3256-70.

25. Song AA-L, In LLA, Lim SHE, Rahim RA. 2017. A review on Lactococcus lactis: from food to factory. Microbial Cell Factories 16:55.

26. Kuipers OP, Beerthuyzen MM, de Ruyter PG, Luesink EJ, de Vos WM. 1995. Autoregulation of nisin biosynthesis in *Lactococcus lactis* by signal transduction. Journal of Biological Chemistry 270:27299-304.

27. Mierau I, Kleerebezem M. 2005. 10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*. Applied Microbiology and Biotechnology 68:705-717.

28. Brautaset T, Lale R, Valla S. 2009. Positively regulated bacterial expression systems. Microbial biotechnology 2:15-30.

29. Monne M, Chan KW, Slotboom DJ, Kunji ER. 2005. Functional expression of eukaryotic membrane proteins in *Lactococcus lactis*. Protein Science 14:3048-56.

30. Leroy F, De Vuyst L. 2004. Lactic acid bacteria as functional starter cultures for the food fermentation industry. Trends in Food Science and Technology 15:67-78.

31. Settanni L, Moschetti G. 2014. New trends in technology and identity of traditional dairy and fermented meat production processes: Preservation of typicality and hygiene. Trends in Food Science and Technology 37:51-58.

32. Pujato SA, Quiberoni A, Mercanti DJ. 2018. Bacteriophages on dairy foods. Journal of Applied Microbiology 126:14-30.

33. Mahony J, Cambillau C, van Sinderen D. 2017. Host recognition by lactic acid bacterial phages. FEMS Microbiology Reviews 41:S16-S26.

34. Mahony J, van Sinderen D. 2014. Current taxonomy of phages infecting lactic acid bacteria. Frontiers in Microbiology 5:7.

35. Deveau H, Labrie SJ, Chopin M-C, Moineau S. 2006. Biodiversity and classification of lactococcal phages. Applied and Environmental Microbiology 72:4338-4346.

36. Oliveira J, Mahony J, Lugli GA, Hanemaaijer L, Kouwen T, Ventura M, van Sinderen D. 2016. Genome sequences of eight prophages isolated from *Lactococcus lactis* dairy strains. Genome Announcements 4:e00906-16.

37. Kot W, Neve H, Vogensen FK, Heller KJ, Sørensen SJ, Hansen LH. 2014. Complete genome sequences of four novel *Lactococcus lactis* phages distantly related to the rare 1706 phage species. Genome Announcements 2:e00265-14.

38. Kelly WJ, Altermann E, Lambie SC, Leahy SC. 2013. Interaction between the genomes of *Lactococcus lactis* and phages of the P335 species. Frontiers in Microbiology 4:257.

39. Mahony J, Randazzo W, Neve H, Settanni L, van Sinderen D. 2015. Lactococcal 949 group phages recognize a carbohydrate receptor on the host cell surface. Applied and Environmental Microbiology 81:3299-3305.

40. Murphy J, Bottacini F, Mahony J, Kelleher P, Neve H, Zomer A, Nauta A, van Sinderen D. 2016. Comparative genomics and functional analysis of the 936 group of lactococcal *Siphoviridae* phages. Scientific Reports 6:21345.

41. Mahony J, Oliveira J, Collins B, Hanemaaijer L, Lugli GA, Neve H, Ventura M, Kouwen TR, Cambillau C, van Sinderen D. 2017. Genetic and functional characterisation of the lactococcal P335 phage-host interactions. BMC Genomics 18:146.

42. Rousseau GM, Moineau S. 2009. Evolution of *Lactococcus lactis* phages within a cheese fsactory. Applied and Environmental Microbiology 75:5336-5344.

43. Murphy J, Royer B, Mahony J, Hoyles L, Heller K, Neve H, Bonestroo M, Nauta A, van Sinderen D. 2013. Biodiversity of lactococcal bacteriophages isolated from 3 Gouda-type cheese-producing plants. Journal of Dairy Science 96:4945-4957.

44. Mahony J, Murphy J, van Sinderen D. 2012. Lactococcal 936-type phages and dairy fermentation problems: from detection to evolution and prevention. Frontiers in Microbiology 3:335.

45. Chmielewska-Jeznach M, Bardowski JK, Szczepankowska AK. 2018. Molecular, physiological and phylogenetic traits of *Lactococcus* 936-type phages from distinct dairy environments. Scientific Reports 8.

46. Oliveira J, Mahony J, Hanemaaijer L, Kouwen TRHM, van Sinderen D. 2018. Biodiversity of bacteriophages infecting *Lactococcus lactis* starter cultures. Journal of Dairy Science 101:96-105.

47. Stuer-Lauridsen B, Janzen T, Schnabl J, Johansen E. 2003. Identification of the host determinant of two prolate-headed phages infecting *Lactococcus lactis*. Virology 309:10-17.

48. Lubbers MW, Waterfield NR, Beresford TP, Le Page RW, Jarvis AW. 1995. Sequencing and analysis of the prolate-headed lactococcal bacteriophage c2 genome and identification of the structural genes. Applied and Environmental Microbiology 61:4348-4356.

49. Labrie SJ, Josephsen J, Neve H, Vogensen FK, Moineau S. 2008. Morphology, genome sequence, and structural proteome of type phage P335 from *Lactococcus lactis*. Applied and Environmental Microbiology 74:4636-4644.

50. Mahony J, Martel B, Tremblay DM, Neve H, Heller KJ, Moineau S, van Sinderen D. 2013. Identification of a new P335 subgroup through molecular analysis of lactococcal phages Q33 and BM13. Applied and Environmental Microbiology 79:4401-4409.

51. Mahony J, Deveau H, Mc Grath S, Ventura M, Canchaya C, Moineau S, Fitzgerald GF, van Sinderen D. 2006. Sequence and comparative genomic analysis of lactococcal bacteriophages jj50, 712 and P008: evolutionary insights into the 936 phage species. FEMS Microbiology Letters 261:253-61.

52. Millen AM, Romero DA. 2016. Genetic determinants of lactococcal C2 viruses for host infection and their role in phage evolution. The Journal of General Virology 97:1998-2007.

53. Kelleher P, Mahony J, Schweinlin K, Neve H, Franz CM, van Sinderen D. 2018. Assessing the functionality and genetic diversity of lactococcal prophages. International Journal of Food Microbiology 272:29-40.

54. Dupont K, Janzen T, Vogensen FK, Josephsen J, Stuer-Lauridsen B. 2004. Identification of *Lactococcus lactis* genes required for bacteriophage adsorption. Applied and Environmental Microbiology 70:5825-32.

55. Ainsworth S, Sadovskaya I, Vinogradov E, Courtin P, Guerardel Y, Mahony J, Grard T, Cambillau C, Chapot-Chartier M-P, van Sinderen D. 2014. Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. mBio 5:e00880-14.

56. Chapot-Chartier MP, Vinogradov E, Sadovskaya I, Andre G, Mistou MY, Trieu-Cuot P, Furlan S, Bidnenko E, Courtin P, Pechoux C, Hols P, Dufrene YF, Kulakauskas S. 2010. Cell surface of *Lactococcus lactis* is covered by a protective polysaccharide pellicle. Journal of Biological Chemistry 285:10464-71.

57. McCabe O, Spinelli S, Farenc C, Labbé M, Tremblay D, Blangy S, Oscarson S, Moineau S, Cambillau C. 2015. The targeted recognition of *Lactococcus lactis* phages to their polysaccharide receptors. Molecular Microbiology 96:875-886.

58. Monteville MR, Ardestani B, Geller BL. 1994. Lactococcal bacteriophages require a host cell wall carbohydrate and a plasma membrane protein for adsorption and ejection of DNA. Applied and Environmental Microbiology 60:3204-11.

59. Collins B, Bebeacua C, Mahony J, Blangy S, Douillard FP, Veesler D, Cambillau C, van Sinderen D. 2013. Structure and functional analysis of the host recognition device of lactococcal phage Tuc2009. Journal of Virology 8429-40.

60. Stockdale SR, Mahony J, Courtin P, Chapot-Chartier MP, van Pijkeren JP, Britton RA, Neve H, Heller KJ, Aideh B, Vogensen FK, van Sinderen D. 2013. The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for infection specialization. Journal of Biological Chemistry 288:5581-90.

61. Sciara G, Blangy S, Siponen M, Mc Grath S, van Sinderen D, Tegoni M, Cambillau C, Campanacci V. 2008. A topological model of the baseplate of lactococcal phage Tuc2009. Journal of Biological Chemistry 283:2716-23.

62. Bebeacua C, Lai L, Vegge CS, Brondsted L, van Heel M, Veesler D, Cambillau C. 2013. Visualizing a complete *Siphoviridae* member by single-particle electron microscopy: the structure of lactococcal phage TP901-1. Journal of Virology 87:1061-8.

63. Veesler D, Spinelli S, Mahony J, Lichiere J, Blangy S, Bricogne G, Legrand P, Ortiz-Lombardia M, Campanacci V, van Sinderen D, Cambillau C. 2012. Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. PNAS 109:8954-8.

64. Chapot-Chartier MP. 2014. Interactions of the cell-wall glycopolymers of lactic acid bacteria with their bacteriophages. Frontiers Microbiology 5:236.

65. Mahony J, Kot W, Murphy J, Ainsworth S, Neve H, Hansen LH, Heller KJ, Sørensen SJ, Hammer K, Cambillau C, Vogensen FK, van Sinderen D. 2013. Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. Applied and Environmental Microbiology 79:4385-4392.

66. Legrand P, Collins B, Blangy S, Murphy J, Spinelli S, Gutierrez C, Richet N, Kellenberger C, Desmyter A, Mahony J, van Sinderen D, Cambillau C. 2016. The atomic structure of the phage Tuc2009 baseplate tripod suggests that host recognition involves two different carbohydrate binding modules. MBio 7:e01781-15.

67. Sciara G, Bebeacua C, Bron P, Tremblay D, Ortiz-Lombardia M, Lichière J, van Heel M, Campanacci V, Moineau S, Cambillau C. 2010. Structure of lactococcal phage p2 baseplate and its mechanism of activation. PNAS 107:6852-6857.

68. Mahony J, Tremblay DM, Labrie SJ, Moineau S, van Sinderen D. 2015. Investigating the requirement for calcium during lactococcal phage infection. International Journal of Food Microbiology 201:47-51.

69. Dunne M, Hupfeld M, Klumpp J, Loessner M. 2018. Molecular basis of bacterial host interactions by gram-positive targeting bacteriophages. Viruses 10:397.

70. Räisänen L, Schubert K, Jaakonsaari T, Alatossava T. 2004. Characterization of lipoteichoic acids as *Lactobacillus delbrueckii* phage receptor components. Journal of Bacteriology 186:5529-5532.

71. Baptista C, Santos MA, Sao-Jose C. 2008. Phage SPP1 reversible adsorption to *Bacillus subtilis* cell wall teichoic acids accelerates virus recognition of membrane receptor YueB. Journal of Bacteriology 190:4989-96.

72. Derkx PM, Janzen T, Sørensen KI, Christensen JE, Stuer-Lauridsen B, Johansen E. 2014. The art of strain improvement of industrial lactic acid bacteria without the use of recombinant DNA technology. Microbial Cell Factories 13:S5.

73. Takumi K, Takeoka A, Kinouchi T, Kawata T. 1985. Solubilization and partial properties of receptor substance for bacteriophage α2 induced from *Clostridium botulinum* Type A 190L. Microbiology and Immunology 29:1185-1195.

74. Gaidelytė A, Cvirkaitė-Krupovic V, Daugelavicius R, Bamford JKH, Bamford DH. 2006. The Entry Mechanism of Membrane-Containing Phage Bam35 Infecting *Bacillus thuringiensis*. Journal of Bacteriology 188:5925-5934.

75. Li X, Koç C, Kühner P, Stierhof Y-D, Krismer B, Enright MC, Penadés JR, Wolz C, Stehle T, Cambillau C, Peschel A, Xia G. 2016. An essential role for the baseplate protein Gp45 in phage adsorption to *Staphylococcus aureus*. Scientific Reports 6:26455.

76. Bielmann R, Habann M, Eugster MR, Lurz R, Calendar R, Klumpp J, Loessner MJ. 2015. Receptor binding proteins of *Listeria monocytogenes* bacteriophages A118 and P35 recognize serovar-specific teichoic acids. Virology 477:110-118.

77. Li X, Gerlach D, Du X, Larsen J, Stegger M, Kühner P, Peschel A, Xia G, Winstel V. 2015. An accessory wall teichoic acid glycosyltransferase protects *Staphylococcus aureus* from the lytic activity of Podoviridae. Scientific Reports 5:17219.

78. Xia G, Corrigan RM, Winstel V, Goerke C, Grundling A, Peschel A. 2011. Wall teichoic acid-dependent adsorption of staphylococcal siphovirus and myovirus. Journal of Bacteriology 193:4006-9.

79. Munsch-Alatossava P, Alatossava T. 2013. The extracellular phage-host interactions involved in the bacteriophage LL-H infection of *Lactobacillus delbrueckii* ssp. *lactis* ATCC 15808. Frontiers in Microbiology 4:1-5.

80. Dieterle M-E, Spinelli S, Sadovskaya I, Piuri M, Cambillau C. 2017. Evolved distal tail carbohydrate binding modules of *Lactobacillus* phage J-1: a novel type of anti-receptor widespread among lactic acid bacteria phages. Molecular Microbiology 104:608-620.

81. Raimondo LM, Lundh NP, Martinez RJ. 1968. Primary adsorption site of phage PBS1: the flagellum of Bacillus subtilis. J Virol 2:256-64.

82. Lovett PS. 1972. PBP1: A flagella specific bacteriophage mediating transduction in *Bacillus pumilus*. Virology 47:743-752.

83. Heller KJ. 1992. Molecular interaction between bacteriophage and the gram-negative cell envelope. Archives of Microbiology 158:235-248.

84. Montag D, Hashemolhosseini S, Henning U. 1990. Receptor-recognizing proteins of T-even type bacteriophages. The receptor-recognizing area of proteins 37 of phages T4 TuIa and TuIb. Journal of Molecular Biology 216:327-34.

85. Heller K, Braun V. 1979. Accelerated adsorption of bacteriophage T5 to *Escherichia coli* F, resulting from reversible tail fiber-lipopolysaccharide binding. Journal of Bacteriology 139:32-38.

86. Heller K, Braun V. 1982. Polymannose O-antigens of *Escherichia coli*, the binding sites for the reversible adsorption of bacteriophage T5+ via the L-shaped tail fibers. Journal of Virology 41:222-227.

87. Valyasevi R, Sandine WE, Geller BL. 1990. The bacteriophage kh receptor of Lactococcus lactis subsp. cremoris KH is the rhamnose of the extracellular wall polysaccharide. Appl Environ Microbiol 56:1882-9.

88. Valyasevi R, Sandine WE, Geller BL. 1994. *Lactococcus lactis* ssp. *lactis* C2 bacteriophage sk1 receptor involving rhamnose and glucose moieties in the cell wall. Journal of Dairy Science 77:1-6.

89. Giraud MF, Naismith JH. 2000. The rhamnose pathway. Current Opinion in Structural Biology 10:687-96.

90. Dong C, Beis K, Giraud MF, Blankenfeldt W, Allard S, Major LL, Kerr ID, Whitfield C, Naismith JH. 2003. A structural perspective on the enzymes that convert dTDP-d-glucose into dTDP-l-rhamnose. Biochemical Society Transactions 31:532-6.

91. Sadovskaya I, Vinogradov E, Courtin P, Armalyte J, Meyrand M, Giaouris E, Palussière S, Furlan S, Péchoux C, Ainsworth S, Mahony J, van Sinderen D, Kulakauskas S,

Guérardel Y, Chapot-Chartier M-P. 2017. Another brick in the wall: a rhamnan polysaccharide trapped inside peptidoglycan of *Lactococcus lactis*. mBio 8.

92. Vinogradov E, Sadovskaya I, Grard T, Murphy J, Mahony J, Chapot-Chartier M-P, van Sinderen D. 2018. Structural studies of the cell wall polysaccharide from *Lactococcus lactis* UC509.9. Carbohydrate Research 461:25-31.

93. Vinogradov E, Sadovskaya I, Courtin P, Kulakauskas S, Grard T, Mahony J, van Sinderen D, Chapot-Chartier M-P. 2018. Determination of the cell wall polysaccharide and teichoic acid structures from *Lactococcus lactis* IL1403. Carbohydrate Research 462:39-44.

94. Farenc C, Spinelli S, Vinogradov E, Tremblay D, Blangy S, Sadovskaya I, Moineau S, Cambillau C. 2014. Molecular insights on the recognition of a *Lactococcus lactis c*ell wall pellicle by the phage 1358 receptor binding protein. Journal of Virology 88:7005-7015.

95. Mistou MY, Sutcliffe IC, van Sorge NM. 2016. Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in gram-positive bacteria. FEMS Microbiology Reviews 40:464-79.

96. Swoboda JG, Campbell J, Meredith TC, Walker S. 2010. Wall teichoic acid function, biosynthesis, and inhibition. ChemBioChem 11:35-45.

97. Yamashita Y, Shibata Y, Nakano Y, Tsuda H, Kido N, Ohta M, Koga T. 1999. A novel gene required for rhamnose-glucose polysaccharide synthesis in *Streptococcus mutans*. Journal of Bacteriology 181:6556-9.

98. Caliot É, Dramsi S, Chapot-Chartier M-P, Courtin P, Kulakauskas S, Péchoux C, Trieu-Cuot P, Mistou M-Y. 2012. Role of the Group B Antigen of *Streptococcus agalactiae*: A Peptidoglycan-Anchored Polysaccharide Involved in Cell Wall Biogenesis. PLOS Pathogens 8:e1002756.

99. Liszewski Zilla M, Chan YGY, Lunderberg JM, Schneewind O, Missiakas D. 2015. LytR-CpsA-Psr enzymes as determinants of *Bacillus anthracis* secondary cell wall polysaccharide assembly. Journal of Bacteriology 197:343-353.

100. Chan YG-Y, Kim HK, Schneewind O, Missiakas D. 2014. The Capsular Polysaccharide of Staphylococcus aureus Is Attached to Peptidoglycan by the LytR-CpsA-Psr (LCP) Family of Enzymes. Journal of Biological Chemistry 289:15680-15690.

101. Kawai Y, Marles-Wright J, Cleverley RM, Emmins R, Ishikawa S, Kuwano M, Heinz N, Bui NK, Hoyland CN, Ogasawara N, Lewis RJ, Vollmer W, Daniel RA, Errington J. 2011. A widespread family of bacterial cell wall assembly proteins. The EMBO Journal 30:4931-4941.

102. Shibata Y, Yamashita Y, Ozaki K, Nakano Y, Koga T. 2002. Expression and characterization of streptococcal rgp genes required for rhamnan synthesis in Escherichia coli. Infection and immunity 70:2891-2898.

103. Brown S, Santa Maria JP, Walker S. 2013. Wall teichoic acids of gram-positive bacteria. Annual review of microbiology 67:1-28.

104. Kramer NE, van Hijum SAFT, Knol J, Kok J, Kuipers OP. 2006. Transcriptome analysis reveals mechanisms by which *Lactococcus lactis* acquires nisin resistance. Antimicrobial agents and chemotherapy 50:1753-1761.

105. Sijtsma L, Wouters JT, Hellingwerf KJ. 1990. Isolation and characterization of lipoteichoic acid, a cell envelope component involved in preventing phage adsorption, from *Lactococcus lactis* subsp. *cremoris* SK110. Journal of bacteriology 172:7126-7130.

106. Endl J, Seidl HP, Fiedler F, Schleifer KH. 1983. Chemical composition and structure of cell wall teichoic acids of staphylococci. Archives of Microbiology 135:215-23.

107. Endl J, Seidl PH, Fiedler F, Schleifer KH. 1984. Determination of cell wall teichoic acid structure of staphylococci by rapid chemical and serological screening methods. Archives of Microbiology 137:272-80.

108. Bron PA, Tomita S, van S, II, Remus DM, Meijerink M, Wels M, Okada S, Wells JM, Kleerebezem M. 2012. *Lactobacillus plantarum* possesses the capability for wall teichoic acid backbone alditol switching. Microbial Cell Factories 11:123.

109. Vinogradov E, Sadovskaya I, Li J, Jabbouri S. 2006. Structural elucidation of the extracellular and cell-wall teichoic acids of *Staphylococcus aureus* MN8m, a biofilm forming strain. Carbohydrate Research 341:738-43.

110. Chapot-Chartier MP, Kulakauskas S. 2014. Cell wall structure and function in lactic acid bacteria. Microbial Cell Factories 13 Suppl 1:S9.

111. Ginsberg C, Zhang YH, Yuan Y, Walker S. 2006. In vitro reconstitution of two essential steps in wall teichoic acid biosynthesis. ACS Chemical Biology 1:25-8.

112. Zhang YH, Ginsberg C, Yuan Y, Walker S. 2006. Acceptor substrate selectivity and kinetic mechanism of *Bacillus subtilis* TagA. Biochemistry 45:10895-904.

113. Pereira MP, Schertzer JW, D'Elia MA, Koteva KP, Hughes DW, Wright GD, Brown ED. 2008. The wall teichoic acid polymerase TagF efficiently synthesizes poly(glycerol phosphate) on the TagB product lipid III. ChemBioChem 9:1385-90.

114. Schertzer JW, Brown ED. 2008. Use of CDP-Glycerol as an alternate acceptor for the teichoic acid polymerase reveals that membrane association regulates polymer length. Journal of Bacteriology 190:6940-6947.

115. Lazarevic V, Karamata D. 1995. The tagGH operon of *Bacillus subtilis* 168 encodes a two-component ABC transporter involved in the metabolism of two wall teichoic acids. Molecular Microbiology 16:345-55.

116. Meredith TC, Swoboda JG, Walker S. 2008. Late-stage polyribitol phosphate wall teichoic acid biosynthesis in *Staphylococcus aureus*. Journal of Bacteriology 190:3046-56.

117. Pereira MP, D'Elia MA, Troczynska J, Brown ED. 2008. Duplication of teichoic acid biosynthetic genes in *Staphylococcus aureus* leads to functionally redundant poly(ribitol phosphate) polymerases. Journal of Bacteriology 190:5642-9.

118. Brown S, Zhang YH, Walker S. 2008. A revised pathway proposed for *Staphylococcus aureus* wall teichoic acid biosynthesis based on in vitro reconstitution of the intracellular steps. Chemistry & Biology 15:12-21.

119. Brown S, Meredith T, Swoboda J, Walker S. 2010. *Staphylococcus aureus* and *Bacillus subtilis* W23 make polyribitol wall teichoic acids using different enzymatic pathways. Chemistry & Biology 17:1101-10.

120. Shiraishi T, Yokota S, Fukiya S, Yokota A. 2016. Structural diversity and biological significance of lipoteichoic acid in Gram-positive bacteria: focusing on beneficial probiotic lactic acid bacteria. Bioscience of Microbiota, Food and Health 35:147-161.

121. Percy MG, Grundling A. 2014. Lipoteichoic acid synthesis and function in gram-positive bacteria. Annual Review of Microbiology 68:81-100.

122. Iwasaki H, Shimada A, Yokoyama K, Ito E. 1989. Structure and glycosylation of lipoteichoic acids in *Bacillus* strains. Journal of Bacteriology 171:424-9.

123. Fischer W. 1988. Physiology of lipoteichoic acids in bacteria, p 233-302. *In* Rose AH, Tempest DW (ed), Advances in Microbial Physiology, vol 29. Academic Press.

124. Koch HU, Fischer W. 1978. Acyldiglucosyldiacylglycerol-containing lipoteichoic acid with a poly(3-O-galabiosyl-2-O-galactosyl-sn-glycero-1-phosphate) chain from *Streptococcus lactis* Kiel 42172. Biochemistry 17:5275-5281.

125. Reid CW, Vinogradov E, Li J, Jarrell HC, Logan SM, Brisson JR. 2012. Structural characterization of surface glycans from *Clostridium difficile*. Carbohydrate Research 354:65-73.

126. Stortz CA, Cherniak R, Jones RG, Treber TD, Reinhardt DJ. 1990. Polysaccharides from *Peptostreptococcus anaerobius* and structure of the species-specific antigen. Carbohydrate Research 207:101-20.

127. Fischer W. 1994. Lipoteichoic acid and lipids in the membrane of *Staphylococcus aureus*. Medical Microbiology and Immunology 183:61-76.

128. Jorasch P, Wolter FP, Zahringer U, Heinz E. 1998. A UDP glucosyltransferase from *Bacillus subtilis* successively transfers up to four glucose residues to 1,2-diacylglycerol: expression of ypfP in *Escherichia coli* and structural analysis of its reaction products. Molecular Microbiology 29:419-30.

129. Uchikawa K, Sekikawa I, Azuma I. 1986. Structural studies on lipoteichoic acids from four *Listeria* strains. Journal of Bacteriology 168:115-22.

130. Brade L, Brade H, Fischer W. 1990. A 28 kDa protein of normal mouse serum binds lipopolysaccharides of gram-negative and lipoteichoic acids of gram-positive bacteria. Microbial Pathogenesis 9:355-62.

131. Fischer W, Mannsfeld T, Hagen G. 1990. On the basic structure of poly (glycerophosphate) lipoteichoic acids. Biochemistry and Cell Biology 68:33-43.

132. Fischer W, Koch HU, Rosel P, Fiedler F. 1980. Alanine ester-containing native lipoteichoic acids do not act as lipoteichoic acid carrier. Isolation, structural and functional characterization. Journal of Biological Chemistry 255:4557-62.

133. Reichmann NT, Grundling A. 2011. Location, synthesis and function of glycolipids and polyglycerolphosphate lipoteichoic acid in Gram-positive bacteria of the phylum *Firmicutes*. FEMS Microbiology Letters 319:97-105.

134. Grundling A, Schneewind O. 2007. Genes required for glycolipid synthesis and lipoteichoic acid anchoring in *Staphylococcus aureus*. Journal of Bacteriology 189:2521-30.

135. Jorasch P, Warnecke DC, Lindner B, Zahringer U, Heinz E. 2000. Novel processive and nonprocessive glycosyltransferases from *Staphylococcus aureus* and *Arabidopsis thaliana* synthesize glycoglycerolipids, glycophospholipids, glycosphingolipids and glycosylsterols. European Journal of Biochemistry 267:3770-83.

136. Gründling A, Schneewind O. 2007. Synthesis of glycerol phosphate lipoteichoic acid in *Staphylococcus aureus*. PNAS 104:8478-8483.

137. Kovács M, Halfmann A, Fedtke I, Heintz M, Peschel A, Vollmer W, Hakenbeck R, Brückner R. 2006. A functional *dlt* operon, encoding proteins required for incorporation of D-alanine in teichoic acids in gram-positive bacteria, confers resistance to cationic antimicrobial peptides in *Streptococcus pneumoniae*. Journal of Bacteriology 188:5797-5805.

138. MacArthur AE, Archibald AR. 1984. Effect of culture pH on the D-alanine ester content of lipoteichoic acid in *Staphylococcus aureus*. Journal of Bacteriology 160:792-793.

139. Fischer W, Rosel P. 1980. The alanine ester substitution of lipoteichoic acid (LTA) in *Staphylococcus aureus*. FEBS Letters 119:224-6.

140. Hurst A, Hughes A, Duckworth M, Baddiley J. 1975. Loss of D-alanine during sublethal heating of *Staphylococcus aureus* S6 and magnesium binding during repair. Journal of General Microbiology 89:277-84.

141. Collins LV, Kristian SA, Weidenmaier C, Faigle M, Van Kessel KP, Van Strijp JA, Gotz F, Neumeister B, Peschel A. 2002. *Staphylococcus aureus* strains lacking D-alanine modifications of teichoic acids are highly susceptible to human neutrophil killing and are virulence attenuated in mice. The Journal of Infectious Diseases 186:214-9.

142. Kramer NE, Hasper HE, van den Bogaard PT, Morath S, de Kruijff B, Hartung T, Smid EJ, Breukink E, Kok J, Kuipers OP. 2008. Increased D-alanylation of lipoteichoic acid and a thickened septum are main determinants in the nisin resistance mechanism of *Lactococcus lactis*. Microbiology 154:1755-62.

143. Giaouris E, Briandet R, Meyrand M, Courtin P, Chapot-Chartier M-P. 2008. Variations in the degree of D-alanylation of teichoic acids in *Lactococcus lactis* alter resistance to cationic antimicrobials but have no effect on bacterial surface hydrophobicity and charge. Applied and Environmental Microbiology 74:4764-4767.

144. Mancuso DJ, Chiu TH. 1982. Biosynthesis of glucosyl monophosphoryl undecaprenol and its role in lipoteichoic acid biosynthesis. Journal of Bacteriology 152:616-25.

145. Yokoyama K, Araki Y, Ito E. 1988. The function of galactosyl phosphorylpolyprenol in biosynthesis of lipoteichoic acid in *Bacillus coagulans*. European Journal of Biochemistry 173:453-8.

146. Allison SE, D'Elia MA, Arar S, Monteiro MA, Brown ED. 2011. Studies of the genetics, function, and kinetic mechanism of TagE, the wall teichoic acid glycosyltransferase in *Bacillus subtilis* 168. Journal of Biological Chemistry 286:23708-16.

147. Brown S, Xia G, Luhachack LG, Campbell J, Meredith TC, Chen C, Winstel V, Gekeler C, Irazoqui JE, Peschel A, Walker S. 2012. Methicillin resistance in *Staphylococcus aureus* requires glycosylated wall teichoic acids. PNAS 109:18909-14.

148. Spears PA, Havell EA, Hamrick TS, Goforth JB, Levine AL, Abraham ST, Heiss C, Azadi P, Orndorff PE. 2016. *Listeria monocytogenes* wall teichoic acid decoration in virulence and cell-to-cell spread. Molecular Microbiology 101:714-30.

149. Mann E, Whitfield C. 2016. A widespread three-component mechanism for the periplasmic modification of bacterial glycoconjugates. Canadian Journal of Chemistry 94:883-893.

150. Kho K, Meredith TC. 2018. Salt-induced stress stimulates a lipoteichoic acid-specific three-component glycosylation system in *Stsaphylococcus aureus*. Journal of Bacteriology 200.

151. Leenhouts KJ, Kok J, Venema G. 1989. Campbell-like integration of heterologous plasmid DNA into the chromosome of Lactococcus lactis subsp. lactis. Appl Environ Microbiol 55:394-400.

152. Maguin E, Duwat P, Hege T, Ehrlich D, Gruss A. 1992. New thermosensitive plasmid for gram-positive bacteria. Journal of bacteriology 174:5633-5638.

153. Xin Y, Guo T, Mu Y, Kong J. 2017. Development of a counterselectable seamless mutagenesis system in lactic acid bacteria. Microbial Cell Factories 16:116.

154. Solem C, Defoor E, Jensen PR, Martinussen J. 2008. Plasmid pCS1966, a new selection/counterselection tool for lactic acid bacterium strain construction based on the oroP gene, encoding an orotate transporter from Lactococcus lactis. Applied and environmental microbiology 74:4772-4775.

155. Ellis HM, Yu D, DiTizio T, Court DL. 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A 98:6742-6.

156. van Pijkeren JP, Britton RA. 2012. High efficiency recombineering in lactic acid bacteria. Nucleic Acids Research 40:e76.

157. Oh JH, van Pijkeren JP. 2014. CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. Nucleic Acids Research 42:e131.

158. Van Pijkeren JP, Neoh KM, Sirias D, Findley AS, Britton RA. 2012. Exploring optimization parameters to increase ssDNA recombineering in *Lactococcus lactis* and *Lactobacillus reuteri*. Bioengineered 3:209-17.

Chapter 2

Identification of genes involved in glucose decoration of the *Lactococcus lactis* cell wall polysaccharide

NB. CWPS structural characterisation and analysis were performed by the group of Dr. Marie-Pierre Chapot-Chartier, Micalis Institute, INRA, Jouy-en-Josas, France, while methylation analysis was performed by Dr. Irina Sadovskaya, University of Littoral Côte d'Opale, Boulogne-sur-mer, France.

Manuscript in Preparation

Table of Contents

2.1 Abstract 67

2.2 Introduction 68

2.3 Materials and Methods 71

2.3.1 Strains and growth conditions 71

2.3.2 Bacteriophage assays 73

2.3.3 Bioinformatic analysis 73

2.3.4 Cloning 74

2.3.5 Recombineering 76

2.3.6 CRISPR-Recombineering 76

2.3.7 Strain fitness 77

2.3.8 Transcriptional analysis 77

2.3.9 Cell wall polysaccharide (CWPS) extraction, purification and analysis 78

2.4 Results 80

2.4.1 Identifying lactococcal homologues of *gtcABC* and *gtlAB* 80

2.4.2 Prevalence of the genes encoding putative sugar-decoration systems 83

2.4.3 CWPS structure of *L. lactis* NZ9000-*csdCD* and overexpressing strains 84

2.4.4 CWPS structure of *L. lactis cflA* negative mutant 89

2.4.5 *csdCD* / *cflA* fitness impact and bacteriophage interactions 91

2.5 Discussion 96

2.6 Acknowledgements 101

2.7 References 101

2.1 Abstract

The molecular mechanism involved in extra-cytoplasmic sugar decoration of glycopolymer components of the cell wall has been described for a number of Gram-positive and -negative bacterial species. This mechanism typically involves a three-component system represented by an undecaprenyl-phosphate (C_{55}-P) sugar-activating glycosyltransferase (C_{55}-P GT), a flippase, and a polytopic glycosyltransferase (PolM GT) dedicated to attaching sugar residues to a specific glycopolymer. In the current study we describe an equivalent system for the decoration of the previously identified phosphopolysaccharide cell wall structure in *Lactococcus lactis*, known as the pellicle (or PSP). Three gene pairs, *csdAB*, *csdCD*, and *csdEF*, which each encode a C_{55}-P GT and a PolM GT, were first identified as potential candidates that may be involved in PSP glycosylation based on sequence similarities. Further genomic comparisons within the *L. lactis* species combined with mutant construction and polysaccharide structural analysis of the products of *csdCD* confirmed their role in the glucosylation of the PSP. A potential flippase-encoding gene was also identified in the *L. lactis* genome. Finally, the PSP decoration by CsdCD was also shown to impact on bacteriophage-host interactions.

2.2 Introduction

Substantial knowledge has been accumulated regarding the molecular and biochemical steps involved in the assembly of various cell envelope constituents of Gram-positive bacteria, such as peptidoglycan (1), lipoteichoic acid (LTA) (2), wall-teichoic acid (WTA) (3), and cell wall polysaccharides (CWPS) (4). Furthermore, some studies have also investigated the molecular mechanism responsible for the attachment of non-essential (sugar) decorations on WTA and CWPS (5, 6). Such decorations have been shown to affect a number of important biological processes in Gram-positive bacteria, including bacteriophage (phage) attachment/infection (6), resistance to antibiotics (7, 8), antigenicity (5), and virulence (8, 9).

A three-component mechanism has previously been proposed to be responsible for the extracytoplasmic addition of sugar residues on bacterial glyco-conjugates, such as lipid A and LPS O-antigens in Gram-negative bacteria, and teichoic acids (TAs) in Gram-positive bacteria (10). This mechanism is characterised by the initial generation of an undecaprenyl-phosphate (C_{55}-P)-sugar intermediate, catalysed by a membrane-anchored glycosyltransferase (GT) which uses a nucleotide-activated sugar substrate. Following this, a so-called flippase is responsible for the re-orientation of the above-mentioned membrane-associated sugar intermediate from the cytoplasmic to the periplasmic/extracellular face of the membrane. The final attachment of the sugar onto the glycoconjugate is commonly catalysed by an integral membrane GT which contains between eight and fourteen transmembrane helices (TMHs) (10). The three functions involved in this process, i.e. C_{55}-P activating GT, flippase and polytopic transmembrane GT (structurally resembling representatives of the GT-C superfamily (11)), have been attributed to specific genes in *Listeria monocytogenes* whose products are involved in the glycosylation of both WTA and LTA moieties (6, 12). GtcA (a flippase enzyme), GtcB (C_{55}-P GT), and GtcC (polytopic transmembrane GT, PolM GT) were shown

to be involved in the galactosylation of WTA in *Ls. monocytogenes* serotype 4nonb (6). Additionally, GtlA (C_{55}-P GT) and GtlB (PolM GT) were shown to attach galactose (Gal) onto the LTA moiety of *Ls. monocytogenes* 10403S (12).

A number of *Lactococcus lactis* strains, including MG1363 (13), 3107 (14) and SMQ-388 (15), produce an apparently unique dual-component CWPS structure: a rhamnose-rich and presumed unexposed polysaccharide chain, known as the rhamnan (16), and a surface-exposed phosphopolysaccharide or so-called polysaccharide pellicle (PSP) (13, 17). The biosynthetic machinery responsible for the production of these two components is encoded by a large (25-30 kb) gene cluster (the *cwps* gene cluster) which encodes proteins responsible for the production of the deoxythimidine diphosphate (dTDP)-L-rhamnose precursor molecule as well as putative rhamnosyltransferases, glycosyltransferases, and several membrane-spanning proteins (17). The importance of this gene cluster in relation to the interaction of *L. lactis* with its environment, and the importance of this glycan for bacteriophage infection, has been defined through studies in which transposon-mediated, spontaneous or directed mutations were identified in this cluster leading to altered phage-host sensitivities (13, 14, 17, 18). Comparative analysis of the *cwps* gene cluster of six *L. lactis* strains revealed a highly conserved set of genes (which includes genes predicted to be responsible for dTDP-L-rhamnose precursor and rhamnan biosynthesis) and a more variable set of genes at the 3' end of a given *cwps* gene cluster (thought to be responsible for the assembly of the PSP) (17).

The PSP subunits exhibit structural diversity between *L. lactis* strains, with a glucose (Glc) side-chain present in the PSP of certain strains, while this substitution may be absent in others. For example, the PSP of both *L. lactis* MG1363 and *L. lactis* SMQ-388, is a polymer of a phosphohexasaccharide repeating unit containing a Glc side-chain (13, 15), while the PSP of *L. lactis* 3107 is composed of a phosphopentasaccharide linear repeating unit (14).

Interestingly, the PSP structure of *L. lactis* SMQ-388 and *L. lactis* 3107 are virtually identical with the exception of the presence of the Glc side-chain in the structure of the former and the replacement of an N-acetylglucosamine (GlcNAc) moiety in the PSP of *L. lactis* SMQ-388 with a Glc moiety in that of *L. lactis* 3107. The genetic divergence of the *cwps* gene cluster of the two strains does not appear to account for the emergence of such a structural discrepancy in their PSP, prompting an investigation into the existence of potential three-component glycosylation mechanisms in *L. lactis* similar to those discovered in *Ls. monocytogenes*.

In the current study, we demonstrate the existence of a gene pair (*csdCD*) whose products exhibit sequence and topological identity to two glycosyltransferases involved in the three-component glycosylation mechanism previously identified in *Ls. monocytogenes* (6, 12). The gene pair products identified in *L. lactis* NZ9000 were shown to be responsible for the attachement of the Glc side-chain to its PSP. Additionally, we observed that these structural modifications increase resistance to bacteriophage predation under particular environmental conditions. Furthermore, we highlighted the existence of two additional homologous gene pairs in the *L. lactis* genome (designated here as *csdAB* and *csdEF*) expanding the potential glycosylation systems found within *L. lactis*, while we also identified a candidate for a flippase-encoding gene (*llnz_02975*, here renamed *cflA*), whose deduced function would complete the three-component glycosylation mechanism.

2.3 Materials and Methods

2.3.1 Strains and growth conditions

Bacterial strains used in this study are listed in Table 1. Strains were grown overnight at 30 ^{o}C in M17 broth and/or on M17 agar (Oxoid Ltd., Hampshire, United Kingdom) supplemented with glucose (5 g /L of M17 medium). Chloramphenicol (Cm^{r}, 5 μg/ml) and erythromycin (Ery^{r}, 5 μg/ml) or tetracycline (Tet^{r}, 10 μg/ml) (Sigma-Aldrich, Missouri, USA) were added to the media where appropriate. For controlled transcription of genes placed under the nisin-inducible promoter, P^{nisA}, nisaplin (DuPont, Copenhagen, Denmark) was supplemented to the growth medium at a final concentration of 40 ng/ml.

Table 1. Strains, plasmids, and phages used in this study

Strain, Plasmid, or Phage	Feature(s)	Reference
Bacterial Strains		
L. lactis subsp. *cremoris* NZ9000	*L. lactis* MG1363 derivative containing *nisRK*, host to phages p2, and sk1	(19)
L. lactis subsp. *cremoris* 3107	Host to phages TP901-1, 66902, and 66903	(20)
L. lactis subsp. *cremoris* NZ9000-*csdC*	NZ9000 with GAATTCG insert in *llnz_03080* (*csdC*) resulting in a TGA stop codon in *csdC*	This work
L. lactis subsp. *cremoris* NZ9000-*csdD*	NZ9000 with GAATTCG insert in *llnz_03075* (*csdD*) resulting in a TGA stop codon in *csdC*	This work
L. lactis subsp. *cremoris* NZ9000-*csdCD*	NZ9000 with GAATTCG insert in *llnz_03075* (*csdD*) and *llnz_03080* (*csdC*) resulting in a TGA stop codon insertion in both genes.	This work
L. lactis subsp. *cremoris* NZ9000-*cflA*	NZ9000 with TAATAGGGG insert in *llnz_02975* (*cflA*) resulting in a TAA and TAG double stop codon in *cflA*	This work
Plasmids		
pJP005	Recombineering-facilitating vector containing *recT*, P^{nisA}, Cm^r	(21)
pNZ44	High-copy expression vector, contains P44 constitutive promotor, Cm^r	(22)
pCNR	Recombineering-facilitating vector containing *recT*, P^{nisA}, Cm^r derived from the low-copy vector pJP005	This work
pVPL3004	Low-copy vector expressing *cas9* along with tracRNA, Ery^r	(23)
pCRISPR	High-copy vector carrying CRISPR repeats and used for inserting target spacer sequences, Tet^r	(23)
pCRISPR::*cflA*	pCRISPR plasmid carrying a CRISPR repeat targeting the recombineered sequence of gene *cflA*, Tet^r	This work
pNZ44::*csdCD*	pNZ44 containing genes *csdC* and *csdD*	This work
pPTPL	*E. coli-L. lactis* promoter-probe vector, Tet^r	(24)
pPTPL::*csdCD-Prom*	pPTPL containing the predicted promoter region of the *csdC* and *csdD* gene pair	This work
Bacteriophages		
p2	936 species, propagated on NZ9000	(25)
sk1	936 species, propagated on NZ9000	(26)
TP901-1	P335 species, propagated on 3107	(20)
66902	936 species, propagated on 3107	(27)

66903	936 species, propagated on 3107	(27)

2.3.2 Bacteriophage assays

Phages used in this study are listed in Table 1. Propagation of phages on their respective host strains was performed as previously described (28). Similarly, both spot/plaque assays (29) and adsorption assays (30) were performed as previously described with sodium chloride added to a final concentration of 1-2 % w/v, where indicated.

2.3.3 Bioinformatic analysis

Candidate genes encoding proteins responsible for glucose side-chain decoration of the PSP of *L. lactis* MG1363 and *L. lactis* SMQ-388 were selected based on amino acid homologies to proteins previously investigated in *Ls. monocytogenes* (6). To this end, the proteins indicated in Table 3 were compared using BLASTP against the compiled coding sequences (CDS) of the *L. lactis* MG1363 genome (Genbank accession number NC_009004.1). Genes in the *L. lactis* MG1363 genome, which are not associated with the *cwps* gene cluster, were selected for further investigation based on their corresponding amino acid similarity to the following in *Ls. monocytogenes* proteins: GtcA, the potential flippase-encoding gene; GtcB and GtlA, as potential C_{55}-P activating GT encoding-genes; and GtcC and GtlB, as potential large transmembrane GT-encoding genes (6, 12). Selected gene sequences from the *L. lactis* MG1363 genome were used to search for homologues in the genomes of *L. lactis* 3107 and *L. lactis* SMQ-388 using BLASTN (Sylvain Moineau, personal communication). Intrinsic properties of protein sequences were assessed using HHPred (31, 32) and TMHMM online predictive tools (33, 34).

Selected genes were analysed based on sequence identity against all publicly available *L. lactis* chromosomal and plasmid sequences, while protein sequence comparisons were performed using all-against-all bi-directional BLAST alignments (35) (cut-off: E-value 0.0001,

with at least 50 % identity across at least 50 % of either protein sequence). A heat-map matrix, with the presence/absence/non-functionality of each of the retrieved candidate genes from all assessed *L. lactis* strains, was created and visualised using MeV suite (V4.9) (36).

2.3.4 Cloning

All recombinant plasmids (Table 1) were generated in *L. lactis* NZ9000 and primers (Table 2), unless otherwise indicated, were ordered from Eurofins MWG (Ebersberg, Germany). The *csdC* and *csdD* genes were amplified using the oligonucleotide pair oIT12-oIT13 with Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Massachusetts, USA) and cloned in tandem in the high copy number, constitutive expression vector pNZ44 (22). The resulting construct was named pNZ44::*csdCD*. Similarly, the upstream genomic region (193 bp) of gene pair *csdCD* was amplified using the oligonucleotide pair oIT14-oIT15 and cloned into the low copy number, promoter-probe vector, pPTPL (24). The resulting construct was named pPTPL::*csdCD*-Prom.

Table 2. Oligonucleotides used in this study.

Oligo name	Sequence (5'-3')	Target/Comment
oIT01	gatatcatcgtgatggattgtggc	Fwd, 500 bp upstream of inserted sequence used for colony PCR check for nonsense mutations in *csdC*
oIT02	ccacgatgtatttccggcaac	Rev, 500 bp downstream of inserted sequence used for colony PCR check for nonsense mutations in *csdC*
oIT03	gttatccgttgCGAATTC	Rev, oligonucleotide containing the inserted sequence used for the PCR check for anticipated nonsense mutations in *csdC*
oIT04	ggcgaagtttatctaactgttcac	Fwd, 500 bp upstream of inserted sequence site used for colony PCR check for nonsense mutations in *csdD*
oIT05	caatacgtatcagtgtaaggaacc	Rev, 500 bp downstream of inserted sequence used for colony PCR check for nonsense mutations in *csdD*
oIT06	gttagacaagtGAATTCG[1,3]	Fwd, oligonucleotide containing the inserted sequence used for colony PCR check for anticipated nonsense mutations in of *csdD*
oIT07	*aaac*aaacaccaaaaataagatatcttaccgttt*g*	Fwd, oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *cflA*
oIT08	*aaaac*aaacggtaagatatcttatttttggtgttt	Rev, oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *cflA*
oIT09	gattcaatgctcaactttg	Fwd, 500 bp upstream of inserted sequence used for colony PCR check for nonsense mutations in *cflA*
oIT10	catgcctctattatgcca	Rev, 500 bp downstream of inserted sequence used for colony PCR check for nonsense mutations in *cflA*
oIT11	aaatccgaaTAATAGGGG	Rev, oligonucleotide containing the inserted sequence used for colony PCR check for anticipated nonsense mutations in *cflA*
oIT12	*aaaaaa*actagttcagttcgactttatattgc	Rev, 3'-end *csdC* gene *L. lactis* for cloning in pNZ44
oIT13	*aaaaaa*ccatgggtataaaaaatgaggagttg[2]	Fwd, 5'-end *csdD* gene *L. lactis* for cloning in pNZ44
oIT14	*aaaaaa*ggatccgcgtctctttggtttgaggcg	Fwd, 193 bp upstream of 5'-end *csdD L. lactis* for cloning in pPTPL
oIT15	*aaaaaa*tctagatataacaactcctcatttttatacc	Rev, 1 bp upstream of 5'-start of *csdD* for cloning in pPTPL
Recombineering		
oIT16	t*a*t*t*a*gataactcttttataacttctaatgtgttatccgttgCGAATTCaaccatcattaataaaaaaatatgttttttctagaggtagt[4]	*csdC L. lactis*
oIT17	c*a*c*a*a*cacctaccatcattgagacaatcaggccattagataCGAATTCacttgtctaaccataatattcttgattgcattgccccaatta	*csdD L. lactis*
oIT18	t*a*c*g*a*catgaaaaaactcatgaatgttttaaaatccgaaTAATAGGGGacggtaagatatcttatttttggtgttttggcaactgctgtc	*cflA L. lactis*

Footnotes:[1]Capital letters: inserted sequence, [2]Italics: Non-genomic nucleotide sequence, [3]Underlined: Restriction site, [4]Asterisk: phosphorothioate modification

2.3.5 Recombineering

Recombineering was performed as previously described (21, 37, 38) with modifications to the assay for use with *L. lactis* NZ9000 on the selected genes: *csdA, csdB¸ csdC, csdD, csdE,* and *csdF*. A total of 500 μg of mutational oligonucleotide (Table 2) containing an in-frame stop codon and containing 5' phosphorothioate linkages, was transformed into *L. lactis* NZ9000 pJP005. Recombineering oligonucleotides were obtained from Integrated DNA Technologies (Leuven, Belgium).

2.3.6 CRISPR-Recombineering

CRISPR-cas9-assisted recombineering was adapted from a previous publication (23) and was used to insert a nonsense mutation in the *cflA* gene. Initial attempts to create a *cflA* mutant using the above-mentioned recombineering protocol proved unsuccessful possibly due to decreased fitness of such mutants. Briefly, a novel plasmid, named pCNR which contains the replication genes *repA, repD, repE* from the backbone of the plasmid pPTPi (24) along with the chloramphenicol resistance gene (cm^r), P^{niAs} promoter, and the ss-DNA binding protein encoding gene, *recT*, from plasmid pJP005 (21), was constructed. Subsequently, the single-step approach to CRISPR-Cas9-assisted recombineering was employed as was previously outlined (24). Competent cells of *L. lactis* NZ9000 carrying both pVPL3004 (expressing Cas9 and tracrRNA) and pCNR (RecT-expression plasmid) were prepared as previously described (21). These cells were co-transformed with both 100 μg of the recombineering oligonucleotide (Table 2) and 100 ng of the pCRISPR::*cflA* construct (Table 1). The latter plasmid allows for the Cas9 protein encoded by the pVPL3004 vector to target specifically the chromosomes that have not incorporated the mutated version of the *cflA* provided by the recombineering oligonucleotide. After recovery, the cells were plated on GM17 plates supplemented with tetracycline (10 μg/ml) and erythromycin (5 μg/ml) selecting for plasmids pCNR and pCRISPR::*cflA*, respectively. Colonies were screened using colony PCR and those yielding the

appropriate disruptive sequence insertion were further purified (Table 2). The recombineering oligonucleotides were ordered from Integrated DNA Technologies (Leuven, Belgium).

2.3.7 Strain fitness

The *cflA* gene mutant strain derived using CRISPR-Cas-assisted recombineering (Table 1) was evaluated to compare its growth behaviour with that of the control strains *L. lactis* NZ9000 pJP005 and *L. lactis* NZ9000 pVPL3004/pCRISPR. Triplicates of each strain assayed were grown overnight in M17 supplemented with 0.5 % w/v glucose and any required antibiotic for the maintenance of the CRISPR-Cas-assisted recombineering plasmids. The following day the triplicate overnights were subcultured to an optical density (OD_{600nm}) of ~ 0.1 in the same media and their growth was followed for 8 hours through hourly optical density measurements at 600 nm.

2.3.8 Transcriptional analysis

The transcriptional pattern of *csdCD* was investigated as a function of growth. For this purpose, *L. lactis* NZ9000 carrying pPTPL::*csdCD-Prom* was grown at 30 ^{o}C in 1.5 x M17 supplemented with 0.5 % glucose and assayed for β-galactosidase activity (39) every 45 or 90 minutes. At each time point, an aliquot (100 -1000 µl) of the culture was removed and washed in Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$ at pH 7). The OD_{600nm} of the washed culture was determined and, depending on the growth stage of the culture, 0.1 to 1 ml of the washed culture was used for the β-galactosidase assay while maintaining the final assay volume (1 ml) constant by supplementing with Z buffer as required. To permeabilise the cell membrane, 0.005 % w/v sodium dodecyl sulfate (SDS) was added as well as 10 % v/v chloroform, with an incubation step of 5 minutes at 30 ^{o}C, after which 200 µl of 4 mg/ml ortho-Nitrophenyl-β-galactoside (ONPG) (Sigma-Aldrich, Missouri, USA) was added. The reaction was allowed to proceed at 30 ^{o}C until sufficient colour had developed, at

which point 500 µl of 1 M sodium carbonate was added to stop the reaction. The solutions were centrifuged for 5 minutes at 17,000 × *g* and the absorbance of the supernatants was measured at OD_{420nm} and OD_{550nm}. β-galactosidase activity of the cultures is represented in Miller units and was calculated using the following formula: 1000 x [(OD_{420nm} - 1.75 x OD_{550nm})] / (T x V x OD_{600nm}), where T represents the total time of the reaction in minutes and V represents the total volume of the culture, in ml, used in the assay.

2.3.9 Cell wall polysaccharide (CWPS) extraction, purification and analysis

CWPS was extracted from cell envelope fractions prepared from *L. lactis* cells and analysed as described previously (16). Briefly, cells obtained from an exponential phase culture (400 ml at OD_{600nm} 0.6-0.8) were successively treated with SDS, proteases, and nucleases, and the resulting insoluble cell walls were then further treated with 48 % hydrofluoric acid (HF) for 48 h at 4 °C. After drying, rhamnan and PSP oligosaccharides released by HF treatment were separated by size exclusion chromatography with a high performance liquid chromatography system (SEC-HPLC) with two columns in tandem (Shodex Sugar KS-804 and KS-803) (16). Elution was performed with Milli-Q water and detection of eluted compounds was performed with a refractometer (2414 Refractive Index Detector, Waters) and/or UV detector at 206 nm. Fractions corresponding to peaks containing rhamnan and PSP oligosaccharides were collected, dried under vacuum and further analysed for composition and mass (16). Monosaccharide composition was determined following trifluoroacetic acid (TFA) hydrolysis by high performance anion exchange chromatography coupled with pulse-amperometric detection (HPAEC-PAD) (ICS5000 system, ThermoFisher scientific). Purified fractions were analysed by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) using 2,5-dihydroxy-benzoic acid (DHB) matrix with an UltrafleXtreme instrument (Bruker Daltonics, Bremen, Germany) (localised on the MetaboHUB platform, Institut Joliot, CEA, Université Paris Saclay, France).

Methylation analysis was performed by the Ciucanu–Kerek procedure (40) modified by Read et al. (1996) (41). Briefly, a PSP oligosaccharide sample was dissolved in 1 ml of dry dimethylsulfoxide (DMSO). Powdered NaOH (about 50 mg) was added and the mixture was stirred for 15 min, then 0.2 ml of methyl iodide was added and the mixture was stirred for 1 h. The reaction was stopped by adding 3 ml of 10 % aqueous $Na_2S_2O_3$. The permethylated product was extracted with $CHCl_3$ (2 mL). The organic phase was washed with water (5 × 2 ml) and evaporated. The product was hydrolysed with 4 M TFA (110 °C, 3 h), dried, reduced with $NaBD_4$, converted into the alditol acetates by conventional methods, and analysed by GC–MS (42). GC was performed on a Trace GC ULTRA system (Thermo Scientific) equipped with a capillary column NMTR5MS (30 m × 0.25 mm) and flame ionization detector (FID) using a temperature gradient of 170 °C (3 min) → 250 °C at 5 °C/min. GC-MS was performed at the PAGéS technical platform (University of Lille) on a Trace GC-Ultra, TSQ quantum GC system (Thermo Scientific) equipped with a capillary column SOLGEL1MS (30 m × 0.25 mm, SGE analytical science), using a temperature gradient of 170 °C → 230 °C at 3 °C/min, followed by a gradient of 230 °C → 20 °C at 10 °C/min.

2.4 Results

2.4.1 Identifying lactococcal homologues of *gtcABC* and *gtlAB*

Structural analysis of the CWPS in certain *L. lactis* strains had previously identified the presence of a Glc side-chain on the PSP produced by *L. lactis* MG1363 (13) and *L. lactis* SMQ-388 (15) (Figure 1). This feature is absent in the PSP isolated from *L. lactis* 3107, which was unexpected given the high sequence identity between the variable region of the *cwps* gene clusters of strains 3107 and SMQ-388 (14) (the three strains also contain the same number of putative glycosyltransferase-encoding genes in their respective *cwps* cluster). This observation suggested that additional genes located beyond the *cwps* gene cluster contribute to the biosynthesis of the final PSP structure. Recently, genes in *Ls. monocytogenes* have been identified that are responsible for the attachment of sugar side-chains on LTA, i.e. *gtlA* and *gtlB* (12), or on WTA, i.e. *gtcA*, *gtcB* and *gtcC* (6). We used the corresponding protein sequences to identify candidate proteins that may be responsible for the Glc substitution of the PSP of *L. lactis* NZ9000. The *Ls. monocytogenes gtcB* and *gtcC* genes encode proteins that contain multiple predicted transmembrane regions (GtcB contains two TMHs in its C-terminal portion, while GtcC is a PolM protein) (6). The *L. lactis* NZ9000 genome contains three gene pairs with similar characteristics: the putative C_{55}-P GT-encoding genes *llnz_00690* (*csdA*), *llnz_03080* (*csdC*), *llnz_07820* (*csdE*), plus their adjacent and putative PolM GT-encoding genes *llnz_00695* (*csdB*), *llnz_03075* (*csdD*), and *llnz_07825* (*csdF*), respectively (Figure 2). At the protein level, the three predicted C_{55}-P GTs encoded by *L. lactis* NZ9000 exhibit varying levels of similarity to each other and to the *Ls. monocytogenes* GtcB (33 % - 72 % amino acid identity), while the putative PolM GTs proteins do not exhibit such a high degree of identity (Figure 2).

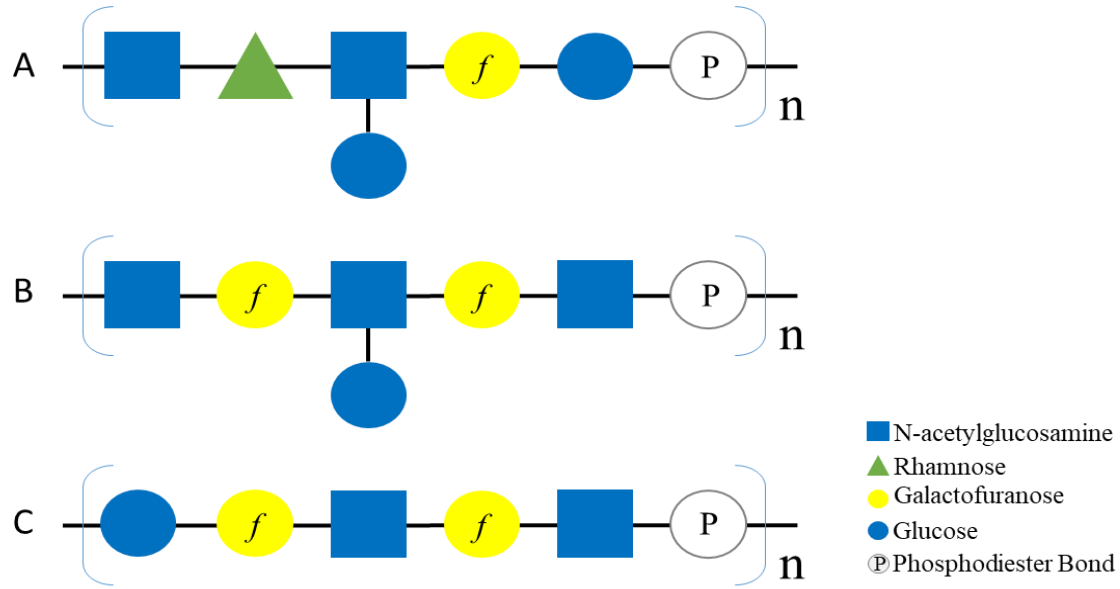


Figure 1. Schematic representation of the PSP structures elucidated in *L. lactis* MG1363 (A), SMQ-388 (B), and 3107 (C).

Sequence comparisons of the three identified *L. lactis* NZ9000 gene pairs with the genomes of *L. lactis* 3107 and SMQ-388 (Figure 3) revealed that only *csdEF* is shared by all three strains. *L. lactis* SMQ-388 also contains homologues of *csdCD*, but appears to lack *csdAB*, while *L. lactis* 3107 lacks both *csdAB* and a functional copy of *csdCD*. The *L. lactis* 3107 genome contains a version of *csdCD* that is presumed to be non-functional due to the presence of a transposase-specifying sequence in *csdC*.

Interestingly, a gene encoding the third protein involved in the glycosylation of *Ls. monocytogenes* WTA (GtcA), which is suggested to encode a flippase, is present in the *L. lactis* NZ9000 genome and corresponds to locus tag *llnz_02975* (designated here as *cflA* [carbohydrate flippase]) and was identified based on protein sequence identity to GtcA. Similar to its equivalent in *Ls. monocytogenes*, the CflA protein is predicted to represent a small (148 amino acid) integral membrane protein containing a total of four TMHs.

Table 3. Proteins involved in the decoration of glycopolymer structures by the addition of sugar side-chains in *Ls. monocytogenes* and their homologues in *L. lactis* NZ9000.

Protein Name	Predicted Protein Function in *Ls. monocytogenes*	*L. lactis* NZ9000 Homologue	% Identity	e-value	Reference
GtcA	Flippase	CflA	26	6e-05	(6, 12)
GtcB / GtlA	C_{55}-P Gal synthase	CsdE	55 / 41	5e-121 / 5e-77	
		CsdA	33 /37	5e-55 / 3e-73	
		CsdC	34 /40	8e-46 / 1e-68	
GtcC / GtlB	C_{55}-P-Gal:WTA/LTA Gal Transferase	CsdF	25 / 32	9e-26 / 0.12	
		CsdD	22 / 32	1e-02 / 0.22	
		CsdB	38 / 25	0.36 / 0.78	

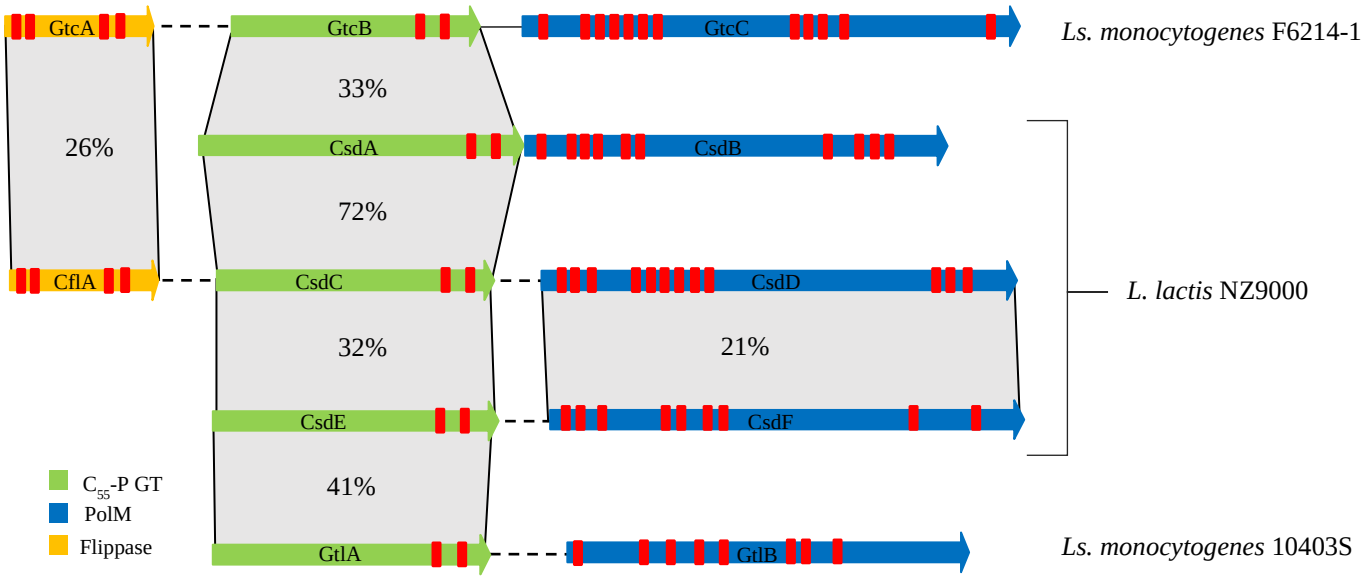


Figure 2. Schematic representation of locus map showing three of the genes, *gtcABC*, associated with WTA galactosylation in *Ls. monocytogenes* F6214-1 and two genes *gtlAB*, associated with LTA galactosylation in *Ls. monocytogenes* 10403S along with seven genes identified in *L. lactis* NZ9000. Genes are colour coded according to predicted functional domains, while percentage similarity based on BLASTP alignment (of the encoded protein products) is also indicated. Finally, sequences representing predicted transmembrane regions (predicted with TMHMM Server v. 2.0) are indicated by red rectangular bands. Solid lines = adjoining genes, Dashed lines = Order/orientation of genes altered for visualization purposes.

2.4.2 Prevalence of the genes encoding putative sugar-decoration systems

A genome wide search of 30 publicly available genomes of *L. lactis*, in addition to *L. lactis* MG1363, 3107, and SMQ-388 failed to identify any additional gene pairs with significant sequence similarity and predicted protein transmembrane regions to those mentioned in Figure 2. Sequence comparisons at the protein level, as seen in Figure 3, indicated the almost ubiquitous presence of *csdEF* in *L. lactis* genomes (29 out of 33 strains), whereas *csdAB* and *csdCD* are much less widely distributed (9 strains and 10 out of a total of 33 examined strains, respectively). A small subset of the genomes appear to contain a functional copy of one of the two genes in a given gene pair whereas the other is predicted to be non-functional due to sequence frameshifts or inserted transposon elements. Furthermore, a protein of high sequence similarity to CflA, with a predicted flippase funtionality, is encoded by all assessed strains. Interestingly, upon further investigation of the genomic region surrounding *csdCD*, it was discovered that in all cases it is either immediately flanked, or interrupted by, transposase-encoding genes or it forms part of a larger predicted (complete or partially complete) prophage genome. Based on these results and in association with the structural data regarding the glucose side-chain found in the PSP of both *L. lactis* MG1363 and *L. lactis* SMQ-388 (and its absence in *L. lactis* 3107), it was hypothesised that the most likely gene candidates responsible for the attachment of the glucose side-chain to PSP would be *csdCD*.

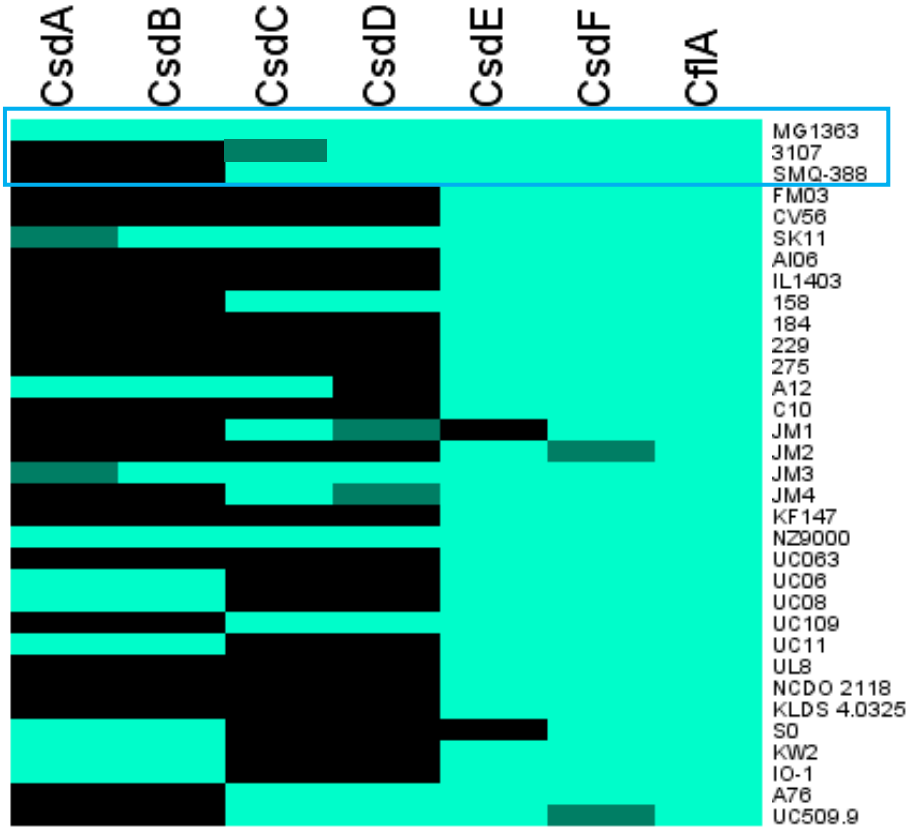


Figure 3. Heatmap representing the presence (light green), absence (black), or non-functionality (dark green) of each of the six proteins encoded by the three glycosyltransferase-associated gene pairs and the flippase (CflA) in a total of 32 *L. lactis* publicly available genomes as well as *L. lactis* SMQ-388 . The heatmap was generated in Mev (V4_9_0). The top three strains, for which structural analysis of the PSP has been performed, are highlighted by a blue frame.

2.4.3 CWPS structure of *L. lactis* NZ9000-*csdCD* and overexpressing strains

To ascertain if the enzymes encoded by *csdCD* contribute to the final PSP structure, several mutants/derivatives of *L. lactis* NZ9000 and 3107 were created. *L. lactis* NZ9000 derivatives carrying non-sense mutations in either *csdC* or *csdD*, or in both genes were made using ssDNA recombineering, resulting in mutant strains *L. lactis* NZ9000-*csdC*, *L. lactis* NZ9000-*csdD*, and *L. lactis* NZ9000-*csdCD*, respectively (Table 1). In parallel, the *csdCD* gene pair was cloned into the high-copy number, high-expression vector pNZ44. The resulting construct was introduced into *L. lactis* 3107, whose PSP structure does not exhibit any glycosyl decorations

and whose genome is lacking a functional homologue of *csdCD* (Figure 3), resulting in strain *L. lactis* 3107 pNZ44::*csdCD* (Table 1).

The composition and structure of the CWPS produced by each of these constructed strains were analysed. The CWPS was extracted from bacterial cell walls by HF treatment and purified by SEC-HPLC allowing separation of rhamnan chains and PSP oligosaccharide subunits as described previously (16). Compositional analysis of the PSP of the *L. lactis* NZ9000-*csdCD* double mutant revealed a clear decrease of Glc/Gal ratio (Glc/Gal ~ 1) compared to that of the PSP of wild-type *L. lactis* NZ9000 (Glc/Gal ~ 2) (Table 4). Structural modification of the PSP from *L. lactis* NZ9000-*csdCD* was confirmed by MALDI-TOF MS analysis. The obtained mass spectrum of the wild-type *L. lactis* NZ9000 purified PSP exhibits a major peak at m/z of 1079.4 corresponding to the hexasaccharide subunit (Figure 4A). This peak was absent in the spectrum of the PSP extracted from the *L. lactis* NZ9000-*csdCD* mutant, which exhibits a major peak at m/z of 917.4, corresponding to a pentasaccharide subunit. This is consistent with the loss of one hexose from the hexasaccharide (Figure 4B), in agreement with the loss of one Glc as revealed by PSP compositional analysis. Of note, secondary peaks at m/z of 914.4 and 755.4 for the *L. lactis* wild-type and NZ9000-*csdCD* mutant PSP, respectively, are presumed to result from HF cleavage immediately following the Gal*f* residue thereby causing the loss of the Glc at the reducing end (Figure 4A and B). Methylation analysis confirmed that the Glc residue absent in the PSP of NZ9000-*csdCD* was the side chain Glc. Indeed, terminal Glc (t-Glc) and 3,6-GlcN (glucosamine) detected in the GC-MS profile of the methylated sugars of the wild-type *L. lactis* NZ9000 PSP disappeared in the *L. lactis* NZ9000-*csdCD* in favour of the appearance of 3-GlcN as expected from the loss of the side-chain Glc branched on the central GlcNAc (data not shown). In order to examine the involvement of the individual *csdC* and *csdD* genes, the structure of PSP purified from the single *csdC* and *csdD* mutants was also analysed. Monosaccharide composition analysis revealed a clear decrease of

Glc/Gal ratio (Glc/Gal ~ 1) in both mutants compared to the PSP of wild-type *L. lactis* NZ9000 (Glc/Gal ~ 2) (Table 4). MALDI-TOF MS analysis and methylation analysis of the PSP obtained from the corresponding *csdD* mutant revealed the complete loss of the side chain Glc, similar to what was observed in the double *csdCD* mutant (Figure 5), indicating the absolute requirement of CsdD for PSP glucosylation. In the case of the *csdC* mutant, MS and methylation analyses indicated that a portion of the PSP (estimated to 20 % based on methylation analysis) still contained the side chain Glc despite *csdC* inactivation. These results suggest that another enzyme with C_{55}-P-Glc synthase activity, possibly CsdA and/or CsdE described above, partially complements *csdC* deficiency by producing the sugar-lipid intermediate further available for CsdD. This hypothesis is further corroborated by the observed amino acid identity between the three C_{55}-P GTs, CsdA, CsdC, and CsdE (Table 5) which indicates that they are proteins of similar function. Altogether, these results demonstrate that the *csdCD* gene pair is involved in Glc side-chain addition on the PSP subunit. Of note, the mass spectrum of the rhamnan purified from *L. lactis* NZ9000-*csdCD* mutant was not modified compared to that of the wild-type (16).

Table 4. Relative monosaccharide composition of PSP purified from various *L. lactis* strains.

Strain	Rha	GlcNAc	Gal[1]	Glc
L. lactis NZ9000	0.8	2.6	1	2.4
L. lactis NZ9000-*csdCD*	0.6	2.1	1	1
L. lactis NZ9000-*csdC*	0.5	2.1	1	1.1
L. lactis NZ9000-*csdD*	0.6	2.5	1	1.1
L. lactis NZ9000-*cflA*	0.7	2	1	1.3
L. lactis 3107 pNZ44	0	1.1	1	0.4
*L. lactis 3*107 pNZ44::*csdCD*	*0.3**	1.0	1	0.9

Footnote: [1]Values are standardised relative to Gal, *Presumed contamination of PSP by rhamnan during SEC-HPLC purification because of shorter rhamnan chains.

Table 5. Percent amino acid identity between the six identified Csd proteins in *Lactococcus lactis.*

	CsdA	**CsdB**	**CsdC**	**CsdD**	**CsdE**	**CsdF**
CsdA	100%	NS[1]	71.57%	NS	33.44%	NH[2]
CsdB	NH	100%	NH	NH	NH	NS
CsdC	71.57%	NH	100%	NH	32.48%	NS
CsdD	NS	NH	NH	100%	NH	20.83%
CsdE	33.44%	NH	32.48%	NH	100%	NS
CsdF	NH	NH	NH	20.83%	NS	100%

Footnote: [1]NS = Not significant, [2]NH = No homology.

The *L. lactis* NZ9000 *csdCD* genes were expressed in *L. lactis* 3107, which itself is devoid of a functional *csdCD* gene pair (Figure 3). Compositional analysis of the PSP purified from *L. lactis* 3107 strain constitutively expressing *csdCD,* revealed an increase of Glc relative to Gal (Glc/Gal ~ 1) compared to the value in the PSP of wild-type 3107 (Glc/Gal ~ 0.4) (Table 4). MALDI-TOF MS analysis confirmed the presence of PSP with an additional hexose in *L. lactis* 3107 pNZ44::*csdCD* with an m/z value of 1095.6 corresponding to an hexasaccharide structure compared to an m/z of 933.4 found for the wild-type PSP pentasaccharide (Figure 4C and D). The other main peaks present on the two spectra correspond to oligosaccharide fragments resulting from HF-mediated cleavage behind the Gal*f* residue, in agreement with the proposed structures (see also above). According to the composition and MS analysis (low abundance of ion with m/z of 933.4), it appears that a high proportion of the PSP subunits of *L. lactis* 3107 pNZ44::*csdCD* strain possess an additional Glc side-chain (Table 4). This was confirmed by methylation analysis which showed the presence of 3,6-GlcN in the PSP of *L. lactis* 3107 pNZ44::*csdCD,* whereas it was absent in the PSP of the control strain *L. lactis* 3107 pNZ44 (data not shown). Regarding the rhamnan, in *L. lactis* 3107 pNZ44::*csdCD*, its chain length appeared to become shorter according to MALDI-TOF MS analysis, with the majority of rhamnan chains being below 3,000 Da in the mutant versus being 5,000 Da in that of wild-type 3107 (data not shown) (16).

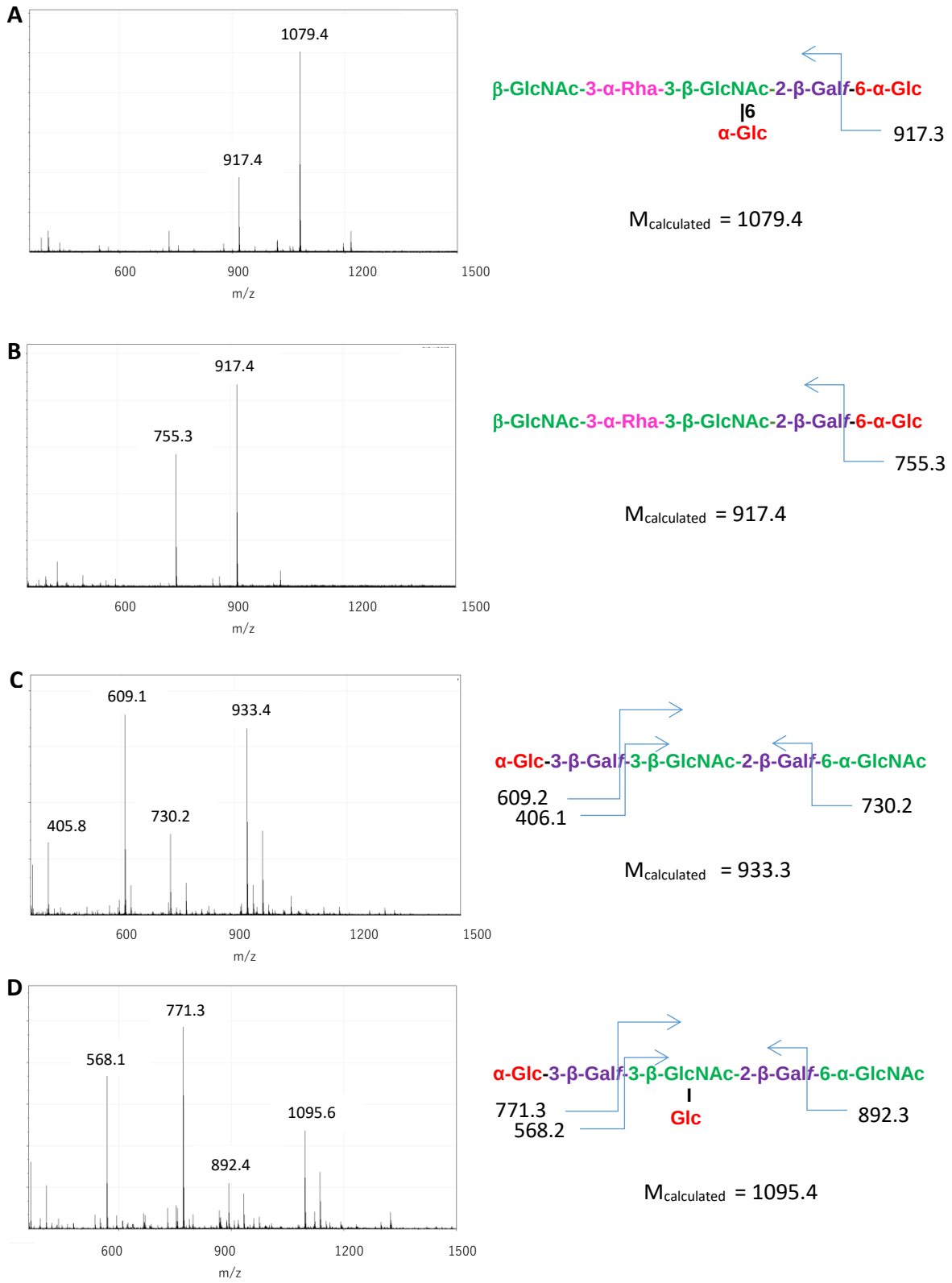


Figure 4. MALDI-TOF MS spectra obtained on purified PSP oligosaccharides extracted by HF treatment (left) and corresponding known or proposed structures with calculated monoisotopic masses ([M+Na]+) (right). m/z values on spectra correspond to [M+Na]+ ion adducts. Arrows indicate the presumed partial cleavage positions caused by HF treatment and figures correspond to calculated monoisotopic masses of derived Na+ adducts. (A) Wild-type *L. lactis* NZ9000; (B) *L. lactis* NZ9000-*csdCD* mutant; (C) wild-type *L. lactis* 3107; (D) *L. lactis* 3107 pNZ44::*csdCD*.

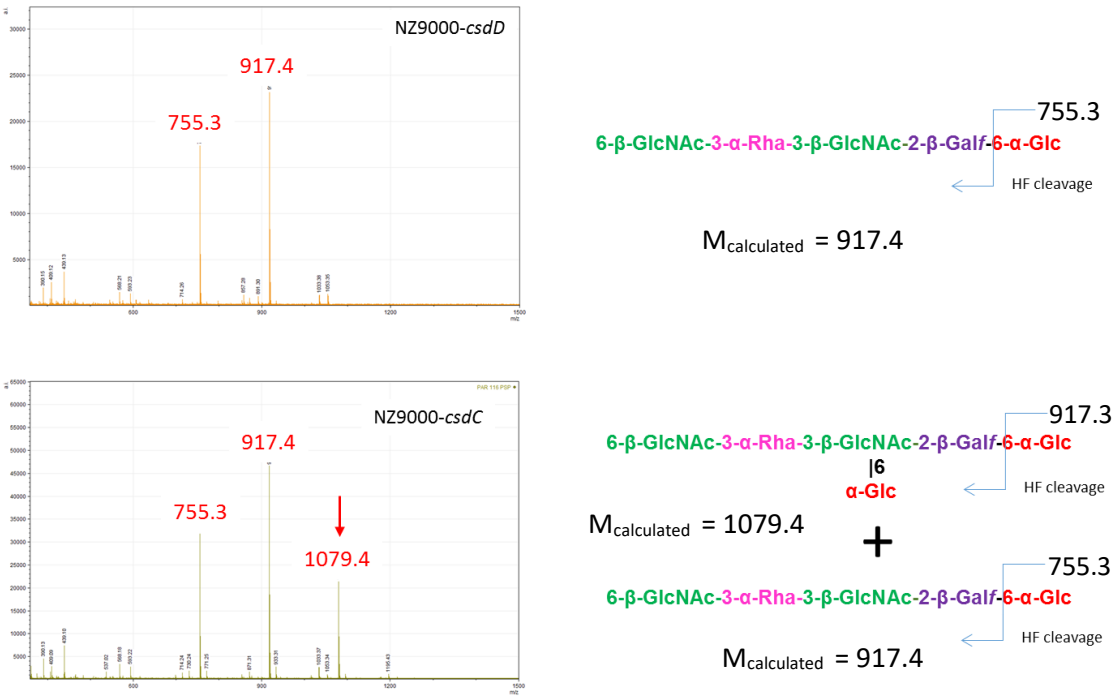


Figure 5. MALDI-TOF MS spectra obtained on purified PSP oligosaccharides extracted from *L. lactis* NZ9000 *csdC* and *csdD* single mutants by HF treatment (left) and corresponding known or proposed structures with calculated monoisotopic masses ([M+Na]+)(right). m/z values on spectra correspond to [M+Na]+ ion adducts. Arrows indicate the presumed partial cleavage positions caused by HF treatment and figures correspond to calculated monoisotopic masses of derived Na+ adducts.

2.4.4 CWPS structure of *L. lactis cflA* negative mutant

As previously proposed (10), a third protein with flippase activity is proposed to be involved in the transfer of the C_{55}-P-Glc intermediate from the inner to the outer side of the cytoplasmic membrane as part of the model three-component glycosylation pathway. By sequence homology search, a gene (*llnz_02975*, here renamed *cflA*) encoding a putative flippase with 26 % amino acid sequence identity to the *Ls. monocytogenes* GtcA flippase (6) and endowed with four putative TMHs as expected, was identified on the *L. lactis* NZ9000 genome (see Table 3 and Figure 3).

In order to investigate the role of CflA in Glc addition on PSP, the *cflA* gene was mutated in *L. lactis* NZ9000. Analysis of the PSP oligosaccharides purified from the *cflA* mutant revealed a decreased level of Glc in PSP purified from the mutant strain compared to the WT, suggesting incomplete addition of Glc side chain onto PSP subunits (Table 4). MS

analysis confirmed this observation with a modified spectrum of the mutant strain versus the WT (Figure 6). The intensities of the peaks with m/z values of 916.71 and 754.45, assigned to a pentasaccharide subunit (devoid of side-chain Glc) and its HF degradation product, were increased relative to the peak at m/z of 1078.90 corresponding to the hexasaccharide subunit. This result strongly suggests that CflA is involved in Glc side chain addition on PSP, yet it also indicates that another, as yet unknown protein is able to flip C55-P-Glc to the outer side of the membrane to be used by CsdD. Of note, we could not identify another putative flippase on the basis of a sequence homology search.

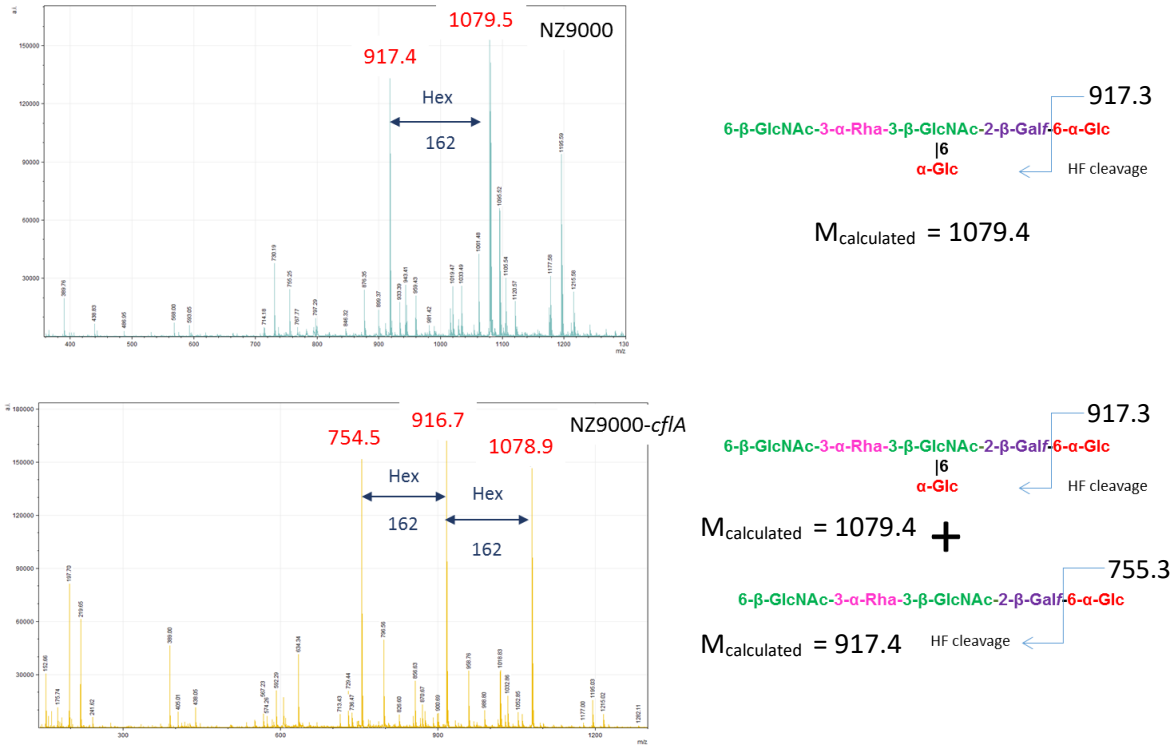


Figure 6. MALDI-TOF MS spectra of PSP oligosaccharides purified by HP-SEC after HF extraction from *L. lactis* NZ9000 WT and NZ9000-*cflA*. m/z values correspond to $[M+Na]^+$ adducts. Corresponding structures are shown on the right. Arrows indicate the presumed HF cleavage sites inside the oligosaccharide structures.

2.4.5 *csdCD* / *cflA* fitness impact and bacteriophage interactions

Lactococcal PSP is required for adsorption by certain phages and mutations that impair or alter PSP biosynthesis may lead to altered phage sensitivities (14, 15, 18). These findings combined with our observation that CsdCD is responsible for the glucosylation of the PSP structure in *L. lactis* NZ9000 prompted us to determine if the PSP modification catalysed by CsdCD may have an impact on phage-host interactions. The mutant strain *L. lactis* NZ9000-*csdCD* was assayed for its sensitivity to the phages listed in Table 1. The presence of the double nonsense mutation did not significantly affect the infective capabilities of the phages under the conditions tested (data not shown). However, a number of altered environmental conditions that more accurately reflect the sub-optimal environments in which phage-host interactions may be taking place, particularly those conditions present in dairy fermentations, were also tested. These included performing the efficiency of plaquing assays in sub-optimal growth temperatures (30 +/- 7 ^{o}C), in a lowered pH environment (pH 4.5-5.5), in the presence of ethanol (1-5 % v/v), and in the presence of salt (up to 2 % w/v). Among these conditions, significantly altered phage sensitivities between *L. lactis* NZ9000 and *L. lactis* NZ9000-*csdCD* were observed in a high salt concentration environment. Figure 7 indicates the much higher efficiency (3.7 log) by which phage p2 can infect the derivative strain *L. lactis* NZ9000-*csdCD* in comparison to the *L. lactis* NZ9000 in the presence of 2 % w/v NaCl, a difference which is not observed under standard laboratory conditions. A similar, but much more subtle effect, can be observed for phage sk1. Similarly, phage adsorption assays revealed that p2 exhibits enhanced reversible and irreversible adsorption to *L. lactis* NZ9000-*csdCD* relative to the wild-type NZ9000, under standard laboratory conditions (Figure 8). No apparent effect on strain fitness was observed for *L. lactis* NZ9000-*csdCD* (data not shown).

In addition to the effect of the *csdCD* double mutants, the effect on bacteriophage interaction was investigated when the flippase gene, *cflA*, alone was mutated. Surprisingly,

despite the effect that the mutation has on the growth of the strain (Figure 9) and the increased sedimentation of the culture relative to the wild-type, the tested phages were equally efficient at infecting the control and mutant *L. lactis* NZ9000-*cflA* strains, both in standard as well as 2 % w/v NaCl conditions (data not shown). This may be explained by the fact that the PSP of *L. lactis* NZ9000-*cflA* still possesses side-chain Glc, albeit at a reduced frequency than the wild-type strain.

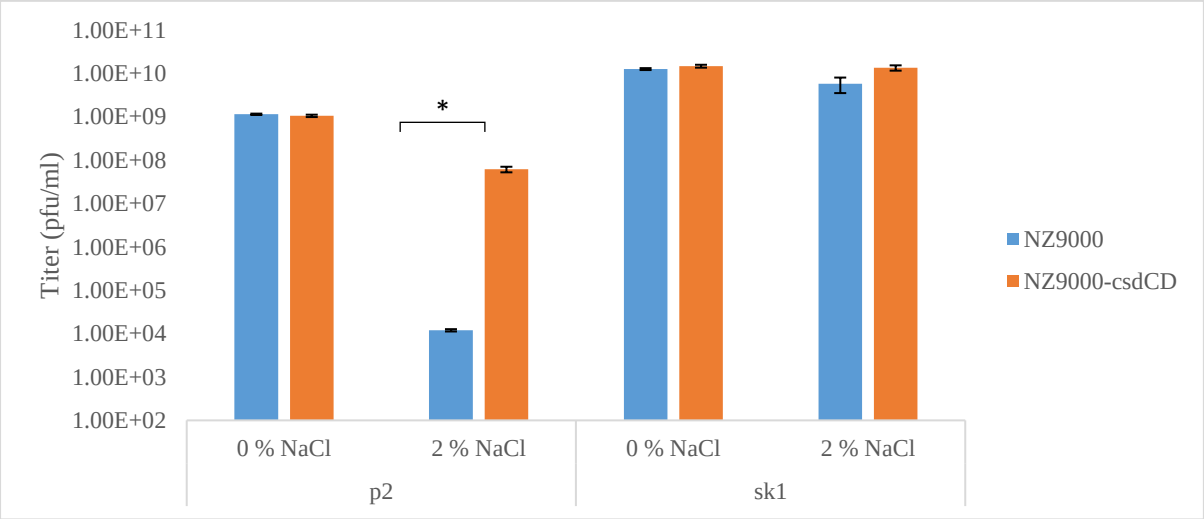


Figure 7. Total infectivity of phages p2 and sk1 against *L. lactis* NZ9000 and *L. lactis* NZ9000-*csdCD* in 0 % and 2 % NaCl environment. p values are indicated by stars *p<0.0001

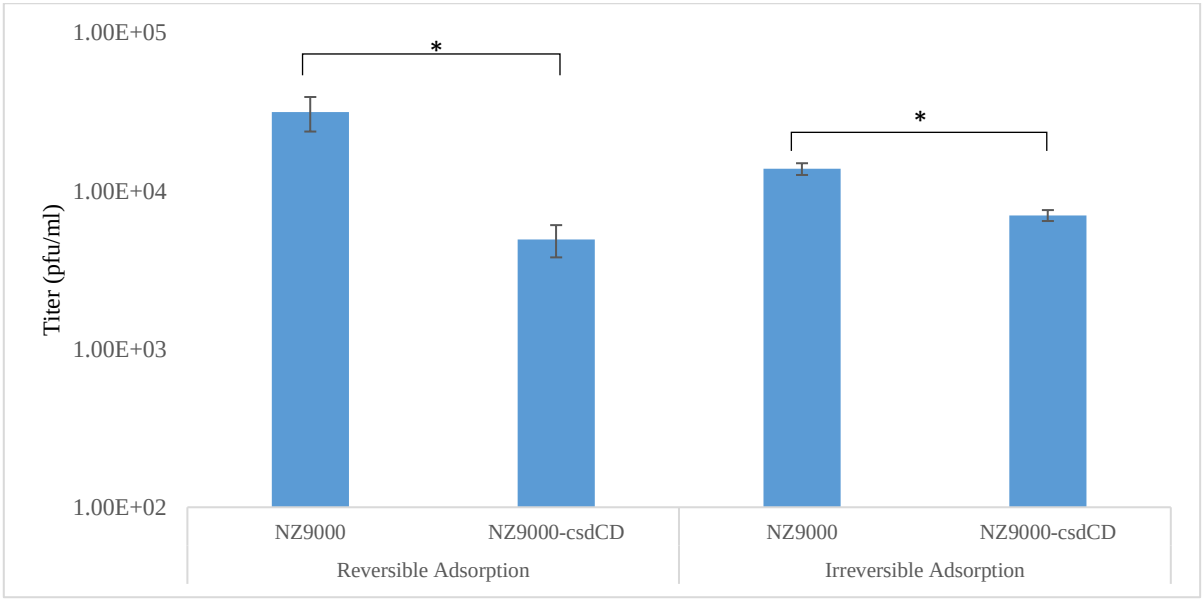


Figure 8. Reversible/Irreversible adsorption assay results indicating the free-phage titres of phage p2 following adsorption to *L. lactis* NZ9000 and *L. lactis* NZ9000-*csdCD*. p values are indicated by stars *p<0.005

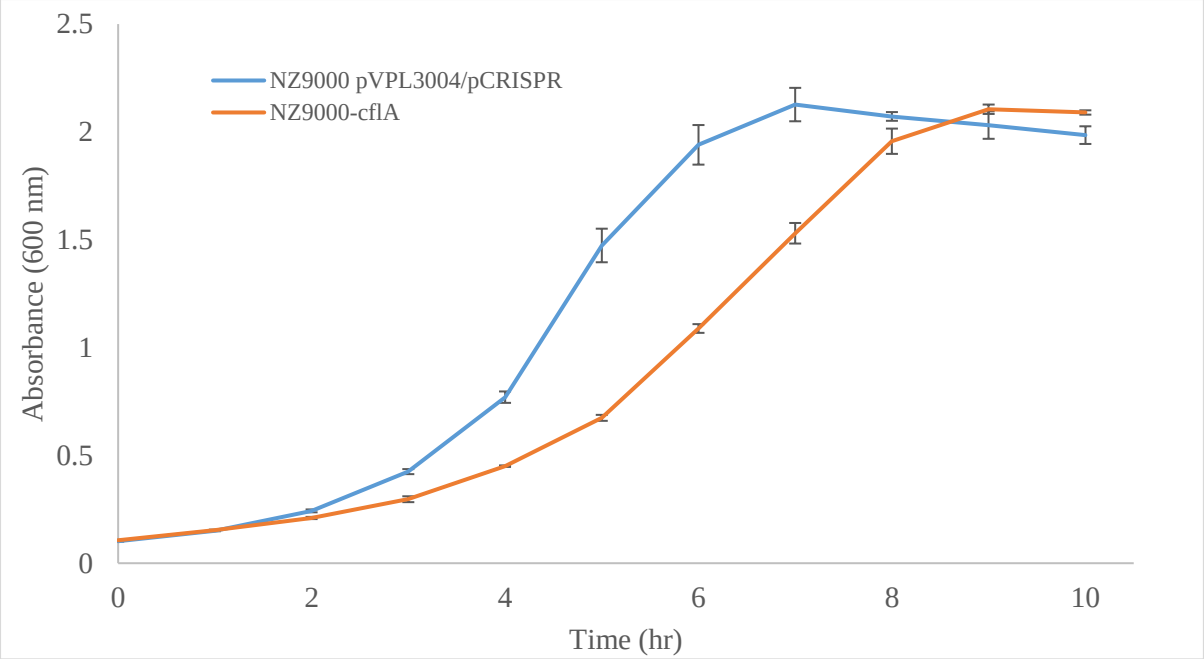


Figure 9. Growth curve of control strain *L. lactis* NZ9000 versus *L. lactis* NZ9000-cflA.

We then tested the impact of overexpressing CsdCD in the strain *L. lactis* 3017 on phage sensitivities. As observed in the case of the *L. lactis* NZ9000-*csdCD* derivative, a subset of the phages tested showed altered infectivity against *L. lactis* 3107 pNZ44::*csdCD* in an increased salt concentration environment (Figure 10). Interestingly, overexpressing *csdCD* in *L. lactis* 3017 seems to have the opposite effect on the infectivity of a P335-type phage representative (TP901-1) and those of 936-type representatives (66902 and 66903). In the case of TP901-1, overexpression of CsdCD inhibited infection in comparison to the wild-type *L. lactis* 3107 in a 1 % w/v NaCl environment, whereas it increased infection capabilities of the 936 phages on the derivative strain in comparison to the wild-type under the same conditions (Note: *L. lactis* 3107 is unable to grow in a medium containing 2 % w/v NaCl) (Figure 10).

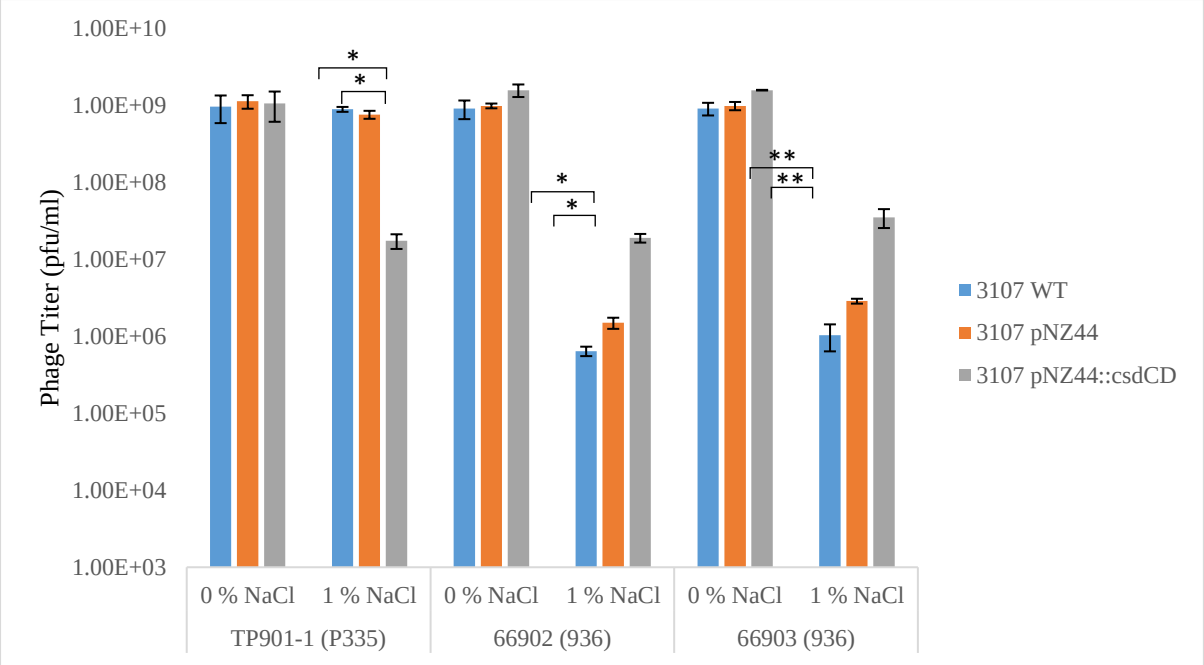


Figure 10. Total infectivity of phages TP901-1, 66902, and 66903 against *L. lactis* 3107 WT. *L. lactis* 3107 pNZ44 and *L. lactis* 3107 pNZ44::*csdCD* in 0 % and 1 % NaCl environment. p values are indicated by stars *p<0.0005, **p<0.005.

2.4.6 Transcriptional analysis of *csdCD*.

In order to assess the transcription of *csdCD*, the complete upstream intergenic sequence of the *csdCD* genes (*csdCD-Prom)* was cloned into the promoter probe vector pPTPL upstream of the *lacZ* gene (Table 1). The derivative plasmid pPTPL::*csdCD-Prom* was transformed into *L. lactis* NZ9000 and the level of transcription was monitored at regular time intervals across a 400-minute time period. As seen in Figure 11, the *lacZ* expression pattern under the control of the *csdCD* promoter region is characteristic of that of a vegetative promoter, closely resembling the growth pattern of the strain and spiking in expression during the exponential phase of the bacterial strain. Interestingly, the growth of *L. lactis* NZ9000 pPTPL appears to be slightly impaired in comparison to that of NZ9000 pPTPL::*csdCD*-Prom. Based the fact that the latter appears to be producing β-galactosidase, which would require additional resources, one would expect the reverse growth phenotypes to be observed. However, the active production of β-

galactosidase, as it is driven by the activity of the *csd*-Prom, appears to be confering a minimal fitness advantage to *L. lactis* NZ9000.

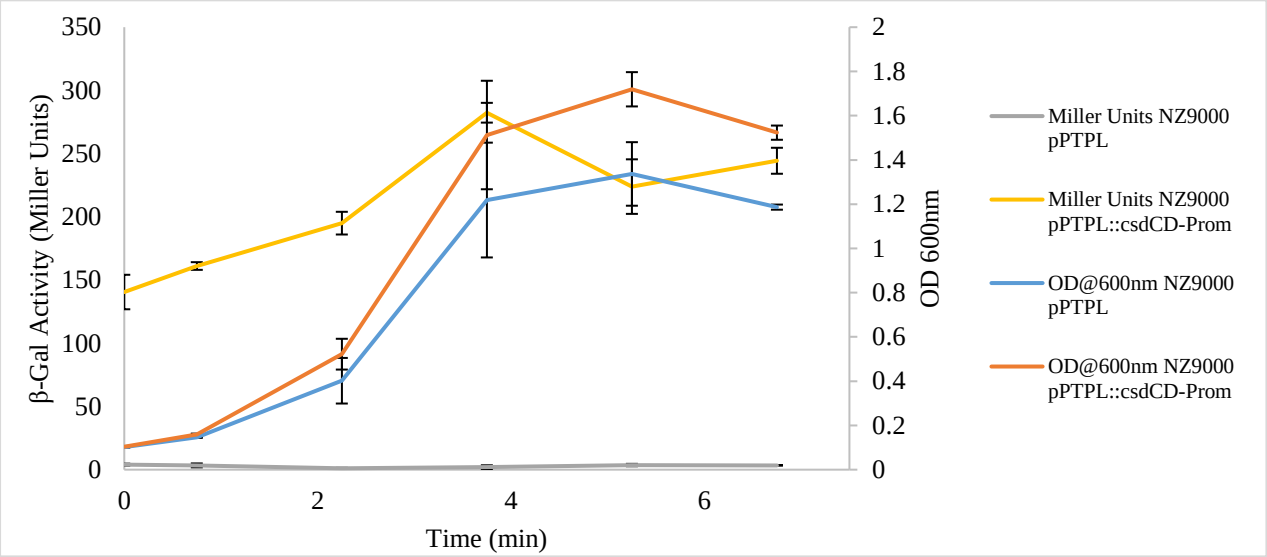


Figure 11. Specific β-Galactosidase activity during growth of the *L. lactis* NZ9000 strain carrying a promoterless pPTPL vector and a pPTPL vector carrying the *csdCD* upstream intergenic region (*csdCD*-Prom).

2.5 Discussion

The genetic machinery responsible for the assembly of the dual-component CWPS structure in *L. lactis* has recently been elucidated for both the rhamnan (16) as well as the PSP (14). The *cwps* gene cluster is predicted to encode eight glycosyltransferases (including rhamnosyltransferases), which in the case of the CWPS structure of *L. lactis* 3107 matches the total number of monosaccharides that make up the subunit structures of each CWPS component (i.e. a trisaccharide rhamnan subunit and a pentasaccharide PSP subunit). In contrast, for *L. lactis* MG1363, a discrepancy exists between the number of glycosyltransferase genes (eight) and the number of sugar moieties in its CWPS (nine). Additionally, it has previously been established, that while the majority of the CWPS biosynthetic steps are encoded by genes that are organised in clusters in both streptococcal and lactococcal species, biosynthetic steps have also been mapped to non-linked genes (4, 5).

In the current work we identified three gene pairs located outside the *cwps* gene cluster, encoding predicted glycosyltransferases (C_{55}-P GT) that show conserved amino acid sequences, and polytopic transmembrane proteins (PolM GT) that appear to be more diverse in amino acid sequence. The *csdEF* gene pair is the most commonly encountered set of the three in *L. lactis,* indicating a generic function. Interestingly, certain strains, including *L. lactis* JM1 and JM2, possess no functional copies of any of the three identified gene pairs. This could be partly explained by the fact that these gene pairs are commonly found in close genomic proximity to other highly mobile elements such as transposases and/or phage components and thus are themselves quite mobile or have been rendered non-operative due to interfering transposases. In addition, the complete absence of any intact *csd* gene hints at the accessory functionality of the gene pair products.

The PSP of *L. lactis* MG1363/NZ9000 is composed of repeating units of a phospho-hexasaccharide containing a Glc side chain. Our results show that the *csdCD* gene pair is

involved in the attachment of the Glc side chain on the PSP subunits. In addition, these genes are also functional when expressed in *L. lactis* 3107 (which lacks a functional homolog of *csdC*), thereby causing Glc addition onto its PSP pentasaccharide linear subunits. CsdC is a putative C_{55}-P glucosyltransferase with two predicted C-terminal TMHs and an intracellular catalytic domain of the GT2 family (Pfam 00535). CsdD is a polytopic transmembrane protein with 12 predicted TMHs and a putative glycosyltranferase domain. Interestingly, these two proteins have the specific structural properties conserved among bacterial proteins that are part of three-component glycosylation systems involved in the extracytoplasmic glycosylation of bacterial glycoconjugates (10). We propose that CsdC and CsdD are part of such a three-component system with CsdC being involved in the transfer of Glc from UDP-Glc to C_{55}-P at the inner face of the cytoplasmic membrane, while CsdD is presumed to catalyse Glc transfer onto PSP outside the cytoplasmic membrane. Interestingly, CsdD is absolutely required for PSP glucosylation, which is in agreement with the presumed specific glucosylation activity on PSP substrate. In contrast, in the absence of CsdC, partial glucosylation of PSP was still observed, indicating that C_{55}-P-Glc substrate can be provided to CsdD by another GT possessing the same activity as CsdC. CsdA and/or CsdE described above that are also predicted to possess C_{55}-P-sugar synthase activity are likely candidates to fulfill this role. Thus, our results indicate a certain level of functional redundancy among different systems dedicated to glycosylation of different cell wall glycopolymers, mediated by the availability of the C_{55}-P sugar intermediate (Table 5). Further work will investigate the substrates of CsdAB and CsdEF in *L. lactis* NZ9000 (see Chapter 3 of this thesis).

As previously described (10), a flippase is proposed to be involved in the transfer of the C_{55}-P-Glc intermediate from the inner to the outer side of the cytoplasmic membrane as part of the three-component glycosylation pathway. A gene encoding a putative flippase with 26 % amino acid sequence identity with the *Ls. monocytogenes* GtcA flippase and with four putative

TMHs as expected, was identified on the *L. lactis* NZ9000 genome. A mutant of *cflA* in *L. lactis* NZ9000 displayed an observable decrease in the total PSP present in the cell wall of the strain as well as decreased Glc substitution of the remaining PSP subunits, strongly indicating that CflA forms part of the three-component glycosylation mechanism in *L. lactis*. Of note, partial substitution of PSP subunits was still present in the *cflA* mutant suggesting the presence of another flippase endowed with the same activity as CflA, although we could not identify it by a sequence homology search. Interestingly, unlike the mutants of *csdCD*, the *cflA* mutant strain showed a distinctly reduced growth rate. Such an effect may result from the indirect of accumulation of C_{55}-P-Glc intermediates preventing recycling of C_{55}-P in order to be reused for other purposes in the cell wall (including peptidoglycan synthesis). It may also be the case that CflA is involved in the glycosylation of other, more essential, glycopolymers, or the growth issues may be caused by a combination of these explanations.

The three step glycosylation mechanism is dedicated to introduce structural modifications late in the biosynthetic pathway thus contributing to the structural diversity of glycoconjugates. It appears to be conserved among bacteria as highlighted recently (10). The prototype systems were described for lipopolysaccharide (LPS) O-antigen glucosylation in the periplasm of Gram-negative bacteria (10), but these systems are also conserved in Gram-positive bacteria and are involved in WTA and LTA glycosylation (6, 12). The *rgpH* and *rgpI* genes (43) which are involved in Glc side-chain formation in the rhamnose-glucose polysaccharide in *Streptococcus pyogenes*, also possess the properties of the GT of the three component systems (RgpH is PolM GT and RgpI exhibits a two TMH profile and is thought to be a C_{55}-P GT). These findings combined with the data presented here extend the range of glycoconjugate substrates of the three component glycosylation mechanisms to the PSP-component of the lactococcal CWPS.

It has previously been established that the receptor for both 936-type phages (17, 18, 44-46) as well as P335-type phages (14, 17, 47) is saccharidic in nature. It was thus expected that modifications of the lactococcal CWPS cause altered phage-host interactions. The structural effect observed when modifying the expression of *csdCD* did not lead to markedly altered phage-host interactions for the majority of tested phages, presumably due to the minor nature of the structural modification of PSP or because the receptor carbohydrate moieties for the assessed phages do not include the central region of the pentasaccharide chain. However, there were representatives of both the 936-type and the P335-type phages whose adsorption efficiency was affected under particular conditions. For example, we observed that phage p2 did not infect the wild-type derivative at an equal efficiency as the structurally altered derivative under saline conditions. Such environmental conditions seemingly destabilise the already weak protein-saccharide interactions characteristic between the 936-type receptor binding proteins (RBPs) and their cell wall polysaccharide receptors. Since phage p2 infects and also adsorbs more efficiently to the *L. lactis* NZ9000 derivative lacking the glucose side-chain, we hypothesize that this phage RBP has evolved to preferentially bind to the primary pentasaccharide unit of the PSP and thus the existence of the glucose-side chain may provide a degree of phage insensitivity to the strain under certain conditions. Similar effects of decoration of cell wall polysaccharide components such as WTA, have previously been reported for a number of Gram positive bacterial strains (3). Such a pronounced effect on infection efficiency is not observed for another representative of the 936-type phages, sk1, despite the high sequence similarity of the RBPs encoded by these two phages. Both phages are predicted to specifically bind to the whole, or part of the, hexasaccharide repeating unit of the *L. lactis* MG1363 PSP (15, 48).

The addition of a glucose-side chain onto the PSP of a closely related strain, such as *L. lactis* 3107, was also shown to cause altered phage-host interactions. This effect is once again

only observed under conditions of high salinity. Both of the 936-type phages tested showed increased infectivity against the derivative strain containing the PSP-associated glucose-side chain in comparison to the control strains. This observation may indicate that there is an evolutionary benefit to *L. lactis* 3107 for not maintaining a functional homolog of *csdCD*, thus decreasing its susceptibility to certain bacteriophages. Finally, the opposite effect on the infectivity of TP901-1 was observed. Under the same assay conditions, TP901-1 showed lower infectivity against *L. lactis* 3107 pNZ44::*csdCD* which indicates that the most effective receptor of TP901-1 is the non-glycosylated version of the strain's PSP.

The transcriptional pattern of *csdCD* closely reflects the growth pattern of the strain, increasing slightly in expression during exponential phase. This expression profile indicates a state of constant flux of the structure of the bacterial cell wall that requires constant addition of the glucose-side chain on any newly formed PSP (1, 13).

The genomic vicinity of *csdCD,* rich in mobile elements including transposases and prophage components, combined with the data regarding phage-host interactions upon modulation of the gene pair expression strongly suggests that such mobile gene cassettes provide a swift response to phage predation by providing non-essential decorations onto an already existing polysaccharide structure, thereby impacting on the infectivity of a phage. It could even be speculated that genes encoding such CWPS modulating functions have been "hijacked" by active prophages in order to further enhance any pre-existing superinfection exclusion mechanisms and activities. Our research reiterates the structural complexity as well the dynamic nature of the CWPS in *L. lactis*. Such characteristics were previously thought to stem from the genomic variability of the central *cwps* gene cluster, yet are now also seen to be influenced from additional genetic components beyond the *cwps* cluster.

2.6 Acknowledgements

Prof. S. Moineau is acknowledged for kindly providing the genomic sequence for the strain *L. lactis* SMQ-388.

2.7 References

1. Chapot-Chartier MP, Kulakauskas S. 2014. Cell wall structure and function in lactic acid bacteria. Microbial Cell Factories 13 Suppl 1:S9.
2. Percy MG, Grundling A. 2014. Lipoteichoic acid synthesis and function in gram-positive bacteria. Annual Review of Microbiology 68:81-100.
3. Brown S, Santa Maria JP, Walker S. 2013. Wall teichoic acids of gram-positive bacteria. Annual review of microbiology 67:1-28.
4. Mistou MY, Sutcliffe IC, van Sorge NM. 2016. Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in gram-positive bacteria. FEMS Microbiology Reviews 40:464-79.
5. Rush JS, Edgar RJ, Deng P, Chen J, Zhu H, van Sorge NM, Morris AJ, Korotkov KV, Korotkova N. 2017. The molecular mechanism of N-acetylglucosamine side-chain attachment to the Lancefield group A Carbohydrate in *Streptococcus pyogenes*. The Journal of Biological Chemistry 292:19441-19457.
6. Spears PA, Havell EA, Hamrick TS, Goforth JB, Levine AL, Abraham ST, Heiss C, Azadi P, Orndorff PE. 2016. *Listeria monocytogenes* wall teichoic acid decoration in virulence and cell-to-cell spread. Molecular Microbiology 101:714-30.
7. Brown S, Xia G, Luhachack LG, Campbell J, Meredith TC, Chen C, Winstel V, Gekeler C, Irazoqui JE, Peschel A, Walker S. 2012. Methicillin resistance in *Staphylococcus aureus* requires glycosylated wall teichoic acids. PNAS 109:18909-14.
8. Carvalho F, Atilano ML, Pombinho R, Covas G, Gallo RL, Filipe SR, Sousa S, Cabanes D. 2015. L-rhamnosylation of *Listeria monocytogenes* wall teichoic acids promotes

resistance to antimicrobial peptides by delaying interaction with the membrane. PLOS Pathogens 11:e1004919.

9. Winstel V, Kühner P, Salomon F, Larsen J, Skov R, Hoffmann W, Peschel A, Weidenmaier C. 2015. Wall teichoic acid glycosylation governs *Staphylococcus aureus* nasal colonization. mBio 6:e00632-15.

10. Mann E, Whitfield C. 2016. A widespread three-component mechanism for the periplasmic modification of bacterial glycoconjugates. Canadian Journal of Chemistry 94:883-893.

11. Gloster TM. 2014. Advances in understanding glycosyltransferases from a structural perspective. Current Opinion in Structural Biology 28:131-141.

12. Rismondo J, Percy MG, Gründling A. 2018. Discovery of genes required for lipoteichoic acid glycosylation predicts two distinct mechanisms for wall teichoic acid glycosylation. The Journal of Biological Chemistry 293:3293-3306.

13. Chapot-Chartier MP, Vinogradov E, Sadovskaya I, Andre G, Mistou MY, Trieu-Cuot P, Furlan S, Bidnenko E, Courtin P, Pechoux C, Hols P, Dufrene YF, Kulakauskas S. 2010. Cell surface of *Lactococcus lactis* is covered by a protective polysaccharide pellicle. Journal of Biological Chemistry 285:10464-71.

14. Ainsworth S, Sadovskaya I, Vinogradov E, Courtin P, Guerardel Y, Mahony J, Grard T, Cambillau C, Chapot-Chartier M-P, van Sinderen D. 2014. Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. mBio 5:e00880-14.

15. Farenc C, Spinelli S, Vinogradov E, Tremblay D, Blangy S, Sadovskaya I, Moineau S, Cambillau C. 2014. Molecular insights on the recognition of a *Lactococcus lactis c*ell wall pellicle by the phage 1358 receptor binding protein. Journal of Virology 88:7005-7015.

16. Sadovskaya I, Vinogradov E, Courtin P, Armalyte J, Meyrand M, Giaouris E, Palussière S, Furlan S, Péchoux C, Ainsworth S, Mahony J, van Sinderen D, Kulakauskas S, Guérardel Y, Chapot-Chartier M-P. 2017. Another brick in the wall: a rhamnan polysaccharide trapped inside peptidoglycan of *Lactococcus lactis*. mBio 8.

17. Mahony J, Cambillau C, van Sinderen D. 2017. Host recognition by lactic acid bacterial phages. FEMS Microbiology Reviews 41:S16-S26.

18. Dupont K, Janzen T, Vogensen FK, Josephsen J, Stuer-Lauridsen B. 2004. Identification of *Lactococcus lactis* genes required for bacteriophage adsorption. Applied and Environmental Microbiology 70:5825-32.

19. Kuipers OP, de Ruyter PGGA, Kleerebezem M, de Vos WM. 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. Journal of Biotechnology 64:15-21.

20. Braun V, Hertwig S, Neve H, Geis A, Teuber M. 1989. Taxonomic differentiation of bacteriophages of *Lactococcus lactis* by electron microscopy, DNA-DNA hybridization, and protein profiles. Microbiology 135:2551-2560.

21. van Pijkeren JP, Britton RA. 2012. High efficiency recombineering in lactic acid bacteria. Nucleic Acids Research 40:e76.

22. McGrath S, Fitzgerald GF, van Sinderen D. 2001. Improvement and optimization of two engineered phage resistance mechanisms in *Lactococcus lactis*. Applied and Environmental Microbiology 67:608-16.

23. Oh JH, van Pijkeren JP. 2014. CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. Nucleic Acids Research 42:e131.

24. O'Driscoll J, Glynn F, Cahalane O, O'Connell-Motherway M, Fitzgerald GF, van Sinderen D. 2004. Lactococcal plasmid pNP40 encodes a novel, temperature-sensitive restriction-modification system. Applied and Environmental Microbiology 70:5546-5556.

25. Higgins DL, Sanozky-Dawes RB, Klaenhammer TR. 1988. Restriction and modification activities from *Streptococcus lactis* ME2 are encoded by a self-transmissible plasmid, pTN20, that forms cointegrates during mobilization of lactose-fermenting ability. Journal of Bacteriology 170:3435-3442.

26. Chandry PS, Moore SC, Boyce JD, Davidson BE, Hillier AJ. 1997. Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. Molecular Microbiology 26:49-64.

27. Oliveira J, Mahony J, Hanemaaijer L, Kouwen TRHM, van Sinderen D. 2018. Biodiversity of bacteriophages infecting *Lactococcus lactis* starter cultures. Journal of Dairy Science 101:96-105.

28. Mahony J, McGrath S, Fitzgerald GF, van Sinderen D. 2008. Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. Applied and Environmental Microbiology 74:6206-6215.

29. Lillehaug D. 1997. An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. Journal of Applied Microbiology 83:85-90.

30. Østergaard Breum S, Neve H, Heller KJ, Vogensen FK. 2007. Temperate phages TP901-1 and φLC3, belonging to the P335 species, apparently use different pathways for DNA injection in *Lactococcus lactis* subsp. *cremoris* 3107. FEMS Microbiology Letters 276:156-164.

31. Soding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research 33:W244-8.

32. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research 33:W244-W248.

33. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology 305:567-80.

34. Sonnhammer EL, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6:175-82.

35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403-10.

36. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34:374-8.

37. Stockdale SR, Collins B, Spinelli S, Douillard FP, Mahony J, Cambillau C, van Sinderen D. 2015. Structure and assembly of TP901-1 virion unveiled by mutagenesis. PLOS ONE 10:e0131676.

38. Van Pijkeren JP, Neoh KM, Sirias D, Findley AS, Britton RA. 2012. Exploring optimization parameters to increase ssDNA recombineering in *Lactococcus lactis* and *Lactobacillus reuteri*. Bioengineered 3:209-17.

39. Israelsen H, Madsen SM, Vrang A, Hansen EB, Johansen E. 1995. Cloning and partial characterization of regulated promoters from *Lactococcus lactis* Tn917-lacZ integrants with the new promoter probe vector, pAK80. Applied and Environmental Microbiology 61:2540-2547.

40. Ciucanu I, & Kerek, F. . 1984. A simple and rapid method for the permethylation ofcarbohydrates. . Carbohydr Res 131:209-217.

41. Read SM, Currie G, Bacic A. 1996. Analysis of the structural heterogeneity of laminarin by electrospray-ionisation-mass spectrometry. Carbohydr Res 281:187-201.

42. Lindberg B, Lonngren J. 1978. Methylation analysis of complex carbohydrates: general procedure and application for sequence analysis. Methods in Enzymology 50:3-33.

43. Ozaki K, Shibata Y, Yamashita Y, Nakano Y, Tsuda H, Koga T. 2002. A novel mechanism for glucose side-chain formation in rhamnose-glucose polysaccharide synthesis. FEBS Letters 532:159-63.

44. Tremblay DM, Tegoni M, Spinelli S, Campanacci V, Blangy S, Huyghe C, Desmyter A, Labrie S, Moineau S, Cambillau C. 2006. Receptor-binding protein of *Lactococcus lactis* phages: identification and characterization of the saccharide receptor-binding site. Journal of Bacteriology 188:2400-10.

45. Geller BL, Ngo HT, Mooney DT, Su P, Dunn N. 2005. Lactococcal 936-species phage attachment to surface of *Lactococcus lactis*. Journal of Dairy Science 88:900-7.

46. Mahony J, Kot W, Murphy J, Ainsworth S, Neve H, Hansen LH, Heller KJ, Sørensen SJ, Hammer K, Cambillau C, Vogensen FK, van Sinderen D. 2013. Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. Applied and Environmental Microbiology 79:4385-4392.

47. Spinelli S, Campanacci V, Blangy S, Moineau S, Tegoni M, Cambillau C. 2006. Modular structure of the receptor binding proteins of *Lactococcus lactis* phages. The RBP structure of the temperate phage TP901-1. Journal of Biological Chemistry 281:14256-62.

48. Spinelli S, Veesler D, Bebeacua C, Cambillau C. 2014. Structures and host-adhesion mechanisms of lactococcal siphophages. Frontiers in Microbiology 5:3.

Chapter 3

Modification of lactococcal cell wall-associated rhamnan and lipoteichoic acids and associated impact on phage infection

NB. CWPS structural characterisation and analysis of lactococcal strains were performed by the group of Dr. Marie-Pierre Chapot-Chartier, Micalis Institute, INRA, Jouy-en-Josas, France. Methylation analysis was performed by Irina Sadovskaya University of Littoral Côte d'Opale, Boulogne-sur-mer, France with the help of the PAGés technical platform (University of Lille).

Manuscript in Preparation

Table of Contents

3.1 Abstract 110

3.2 Introduction 111

3.3 Material and Methods 114

3.3.1 Bacterial strains, phages and growth conditions 114

3.3.2 CWPS and LTA extraction, purification and analysis 116

3.3.3 Bacteriophage assays 118

3.3.4 Cloning 118

3.3.5 Recombineering 120

3.3.6 Transcriptional analysis 120

3.3.7 Strain fitness, survivability, and biofilm formation 120

3.4 Results 122

3.4.1 Identification of *csdAB* and *csdEF* 122

3.4.2 CWPS structure analysis of *L. lactis* NZ9000 *csdAB* mutants and overexpressing strains 122

3.4.3 Expression of *csdAB* genes in *L. lactis* 3107 and CWPS structure analysis 125

3.4.4 CWPS structure of *L. lactis cflA* negative mutants 126

3.4.5 CWPS structure analysis of *L. lactis* NZ9000-*csdEF* and overexpressing strains 131

3.4.6 Role of *csdAB* and *csdEF* in bacteriophage infection 133

3.4.7 Effect of modulating expression of *csdAB* on bacteriophage adsorption 135

3.4.8 Transcriptional analysis of *csdAB* and *csdEF* 136

3.4.9 Strain fitness and biofilm formation with *csdAB* and *csdEF* modulated expression 139

3.5 Discussion 142

3.6 References.....148

3.1 Abstract

Components of the *Lactococcus lactis* cell wall include both cell wall-associated polysaccharides (CWPS) as well as teichoic acids (TA) both of which have been implicated in normal growth of the bacterial cell. Previous studies have identified and structurally characterised both molecules in representative strains of *L. lactis*. From these studies, a degree of structural diversity has been observed between different lactococcal strains. In the case of the CWPS structure, some of these differences have been attributed to genetic diversity within the *cwps* gene cluster. As it was shown in Chapter 2 of this thesis, an additional glycosylation system, encoded by genes *csdCD*, was identified outside of the *cwps* gene cluster. This system was shown to glucosylate a component of the CWPS, known as the polysaccharide pellicle (PSP). Based on the homology of *csdC* and *csdD* to other genes within the *L. lactis* genome, we describe here two additional glucosylation systems encoded by genes *csdAB* and *csdEF*, respectively. The first gene pair was shown to glucosylate the rhamnan component of the CWPS while the latter was shown to be involved in the galactosylation of the lipoteichoic acids (LTA) of the lactococcal cell envelope. Finally, the phenotypic importance of the glycosylation of either structures was explored through the perspective of bacteriophage-host interactions as well as the contribution of these structural modifications to bacterial strain fitness.

3.2 Introduction

Lactococcus lactis is one of the most widely employed species of lactic acid bacteria (LAB) in dairy fermentations. This extensive industrial application has prompted significant scientific interest in its fermentative, organoleptic, and biotechnological abilities (1, 2). Part of this research has focused on the importance and functionality of the lactococcal cell envelope, which in addition to the cytoplasmic membrane is comprised of a thick layer of peptidoglycan (PG) interlinked with other glycopolymers, such as lipoteichoic acids (LTA), wall teichoic acids (WTA), various polysaccharides (PS), and/or (glyco)proteins (3). The PG and its associated polymers are essential for maintaining the bacterial cell shape and integrity throughout its life cycle, particularly during active cell growth and division, while simultaneously mediating interactions with other biotic/abiotic surfaces, due to their exposed location (3-5).

Gram-positive bacteria produce a diverse array of PS including: (i) capsular polysaccharides (CPS), which are commonly covalently linked to the PG and form a thick outer layer, (ii) exopolysaccharides (EPS), which are more loosely associated with the PG and are usually released into the environment, and (iii) cell wall-associated polysaccharides (CWPS), which do not form a capsule and are tightly and probably covalently linked to the PG (3). A number of *L. lactis* strains, including MG1363 (6), 3107 (7), and SMQ-388 (8) have been shown to produce a dual-component CWPS structure. The two components of the structure include a rhamnose-rich and presumed unexposed (i.e. embedded within the PG) polysaccharide chain, known as the rhamnan (9), and a surface-exposed phosphopolysaccharide pellicle (PSP) (6) (10). The machinery responsible for the biosynthesis of these components is encoded by a large gene cluster of 25-30 kbp (the *cwps* gene cluster). Genes within this operon encode proteins involved in the synthesis of the rhamnose precursor molecule as well as putative rhamnosyltransferases, glycosyltransferases, and a number of

membrane-spanning proteins (10). Random and directed mutagenesis of genes within this operon have demonstrated that this gene cluster is required for the interaction of *L. lactis* with its infecting bacteriophages (phages) (6, 7, 10, 11). Comparative analysis of the *cwps* gene clusters of six *L. lactis* strains revealed a highly conserved set of genes (which include those responsible for rhamnose precursor biosynthesis and genes encoding putative rhamnosyltransferases) at the 5' end and a more variable set of genes at the 3' end of the operon, predicted to encode a set of glycosyltransferases as well as a number of glycopolymer assembly enzymes (10). Based on the genetic diversity of the *cwps* gene cluster, the sequenced strains have been classified into three CWPS-specific genotypes (*cwps* A, B, and C) while also based on available genetic diversity more genotypes are expected to exist (12). Furthermore, five unique sub-types (C_1-C_5) of the *cwps* C genotype strains were subsequently identified (7). Biochemical analysis of the PSP of lactococcal strains representing different genotypes (A, B or C) confirmed that the observed genetic variation corresponds to specific PSP structures (6-8, 13, 14).

In addition to the *cwps* gene cluster, recent work has identified a set of genes whose products contribute to the PSP structure of an *L. lactis* strain (see Chapter 2 of this thesis). Specifically, the products of two adjacent genes, *csdC* (*llnz_03080*) and *csdD* (*llnz_03075*) (together referred to as *csdCD*), along with putative flippase *cflA* (*llnz_02975*), were shown to be involved in the decoration of the *L. lactis* NZ9000 PSP structure with a glucose (Glc) side-chain. Furthermore, this Glc substitution ability was shown to be transferable to another CWPS C-type strain, *L. lactis* 3107, when the *csdCD* genes are introduced in strain 3107 on a plasmid. Strains carrying mutations in the two genes lacked the Glc-substitution and caused a moderate, yet statistically significant change, in the mutated strain's sensitivity to certain phages. These results highlight the functional importance of such CWPS modifications, which may be

classified among the several defence mechanisms that bacterial species employ to avoid phage predation.

As mentioned in Chapter 2 of this thesis, the proteins encoded by *csdCD* share their predicted membrane-associated topology and/or display sequence similarity with the products of two additional gene pairs, comprised of a predicted C_{55}-P glycosyltransferase-encoding gene, *csdA* (*llnz_00690*) and *csdE* (*llnz_07820*), and a predicted transmembrane-protein encoding gene, *csdB* (*llnz_00695*) and *csdF* (*llnz_07825*), respectively. In the current study, we demonstrate the functionality of these two gene pairs by highlighting their involvement in the sugar side-chain decoration of two different lactococcal cell wall components, i.e. rhamnan and LTA. We also show the involvement of the candidate flippase (encoded by *llnz_02975*, here renamed *cflA*), in rhamnan glucosylation thus indicating a three-component glycosylation mechanism.

3.3 Material and Methods

3.3.1 Bacterial strains, phages and growth conditions

Bacterial strains, plasmids, and phages used in this study are listed in Table 1. Strains were grown at 30 ^{o}C overnight in M17 broth and/or M17 agar (Oxoid Ltd., Hampshire, United Kingdom) supplemented with 5 g/L glucose. Chloramphenicol (Cm^{r}), tetracycline (Tet^{r}), and streptomycin (Str^{r}) (Sigma-Aldrich, Missouri, USA) were added to the media to a final concentration of 5 μg/ml, 10 μg/ml, and 100 μg/ml respectively, where appropriate. Induction of genes placed under the nisin-controlled promoter, P^{nisA}, was achieved through supplementation of the media with nisaplin (DuPont, Copenhagen, Denmark) at a final concentration of 40 ng/ml.

Table 1. Strains, plasmids and phages used in this study

Strain, plasmid, or phage	Feature(s)	Reference
Bacterial Strains		
L. lactis subsp. *cremoris* NZ9000	*L. lactis* MG1363 derivative containing *nisRK*, host to jj50, p2, and sk1	(15)
L. lactis subsp. *cremoris* 3107	Host to LC3, TP901-1, 66901, 66902, 66903, 62604, 62605, 62601	(16)
L. lactis subsp. *cremoris* VES5751	*L. lactis* MG1363 derivative exhibiting a deficient PSP phenotype due to a spontaneous mutation in *llmg_0226*	(6)
L. lactis subsp. *cremoris* NZ9000-*csdAB*	NZ9000 with GAATTCG insert in *llnz_00690* (*csdA*) and *llnz_00695* (*csdB*), resulting in a TGA stop codon insertion in each gene.	This work
L. lactis subsp. *cremoris* NZ9000-*csdEF*	NZ9000 with GATAACCC insert in *llnz_07820* (*csdE*) and *llnz_07825* (*csdF*), resulting in a TGA and TAA stop codon insertion, respectively.	This work
L. lactis subsp. *cremoris* NZ9000-*cflA*	NZ9000 with TAATAGGGG insert in *llnz_02975* (*cflA*) resulting in a TAA and TAG double stop codon in *cflA*	This work
Plasmids		
pJP005	Recombineering-facilitating vector containing P^{nisA} and associated nisin-inducible *recT*, Cm^r	(17)
pNZ44	High-copy expression vector, contains constitutive P44 promoter , Cm^r	(18)
pNZ44str	Derivative vector of pNZ44, Str^r	Developed by Andrea Garzon, UCC Collection
pNZ44::*csdA*	pNZ44 containing gene *csdA*	This work
pNZ44::*csdB*	pNZ44 containing gene *csdB*	This work
pNZ44::*csdAB*	pNZ44 containing genes *csdA* and *csdB*	This work
pNZ44str::*csdAB*	pNZ44str containing genes *csdA* and *csdB*	This work
pNZ44::*csdE*	pNZ44 containing gene *csdE*	This work
pNZ44::*csdF*	pNZ44 containing gene *csdF*	This work
pNZ44::*csdEF*	pNZ44 containing genes *csdE* and *csdF*	This work
pPTPL	*E. coli-L. lactis* promoter-probe vector, Tet^r	(19)
pPTPL::*csdAB-Prom*	pPTPL containing the predicted promoter region of the *csdA* and *csdB* gene pair	This work
pPTPL::*csdEF-Prom*	pPTPL containing the predicted promoter region of the *csdE* and *csdF* gene pair	This work
Bacteriophages		
jj50*	Spontaneously acquired derivative of jj50 (NC_008371.1), propagated on NZ9000	This work
p2	936 species, propagated on NZ9000	(20)
sk1	936 species, propagated on NZ9000	(21)

MCC1	Derivative of sk1, propagated on NZ9000	UCC Culture Collection
LC3	P335 species, propagated on 3107	(22)
TP901-1	P335 species, propagated on 3107	(16)
66901	936 species, propagated on 3107	(23)
66902	936 species, propagated on 3107	(23)
66903	936 species, propagated on 3107	(23)
62604	936 species, propagated on 3107	(23)
62605	936 species, propagated on 3107	(23)
62601	936 species, propagated on 3107	(23)
63301	P335 species, propagated on 3107	(24)
50101	P335 species, propagated on 3107	(24)
07501	P335 species, propagated on 3107	(25)
58601	P335 species, propagated on 3107	(25)
86501	P335 species, propagated on 3107	(24)

3.3.2 CWPS and LTA extraction, purification and analysis

CWPS was extracted from cell wall fractions of *L. lactis* strains and analysed as described previously (9). Briefly, cells obtained from an exponential phase culture (OD_{600nm} 0.6-0.8) were treated successively with SDS, proteases, nucleases and the remaining insoluble material (cell walls) was further treated with 48 % hydrofluoric acid (HF) for 48 h at 4 °C. After removal of HF under a stream of nitrogen, rhamnan and PSP oligosaccharides released by HF treatment were separated by size exclusion chromatography with an HPLC system (SEC-HPLC) with two columns in tandem (Shodex Sugar KS-804 and KS-803). Elution was performed with Milli-Q H_2O and detection of eluted compounds was performed with a refractometer (2414 Refractive Index Detector, Waters) and/or UV detector at 206 nm. Fractions corresponding to peaks containing rhamnan and PSP oligosaccharides were collected, dried under vacuum and further analysed for saccharidic composition and mass. Monosaccharide composition was determined after trifluoroacetic acid (TFA) hydrolysis, by high performance anion exchange chromatography coupled with pulse-amperometric detection (HPAEC-PAD) (ICS5000 system, Thermo Fisher scientific). Purified fractions were analysed by matrix-assisted laser desorption ionization-time of flight mass spectrometry (Maldi-TOF MS) with 2,5-dihydroxy-benzoic acid

(DHB) matrix with an UltrafleXtreme instrument (Bruker Daltonics, Bremen, Germany) (located at the MetaboHUB platform, Institut Joliot, CEA, Université Paris Saclay, France).

Crude LTA was extracted essentially as described by Sijtsma *et al* (1990) (26). Briefly, cell pellets from 2 liters of culture were defatted by stirring overnight with 50 ml of chloroform-methanol (2:1, v/v). After centrifugation, the defatted cells were dried under the fume-hood and extracted with the mixture 80 % aqueous phenol / 0.1 M sodium acetate buffer, pH 5 at 65 °C with intense stirring for 1 h. The mixture was transferred into Teflon centrifuge tubes and cooled in ice. Phases were separated by centrifugation (4000 x *g*, 30 min). The aqueous phase was collected, deproteinated by addition of trichloroacetic acid (TCA) (up to 5 %), dialyzed extensively against de-ionized water and freeze-dried to give crude LTA extract. Identical amounts (2-5 mg) of LTA extracts of the WT strain *L. lactis* NZ9000 and mutant *L. lactis* NZ9000-*csdEF* were treated with HF (48 %, 4 °C, 24 h) and subjected to comparative analysis for glycerol and monosaccharides, as well as methylation analysis. m-Inositol (Ino) was used as internal standard to assess glycerol content.

Methylation analysis was performed by the Ciucanu–Kerek procedure (27) modified by Read *et al.* (1996) (28). Briefly, a PSP oligosaccharide sample was dissolved in 1 mL of dry DMSO. Powdered sodium hydroxide (NaOH) (about 50 mg) was added and the mixture was stirred for 15 min, then 0.2 mL of methyl iodide (MeI) was added and the mixture was stirred for 1 h. The reaction was stopped by adding 3 mL of 10 % aqueous sodium thiosulfate ($Na_2S_2O_3$). The permethylated product was extracted with chloroform (2 mL). The organic phase was washed with water (5 × 2 mL) and evaporated. The product was hydrolysed with 4 M trifluoroacetic acid TFA (110 °C, 3 h), dried, reduced with sodium borodeuteride ($NaBD_4$), converted to the alditol acetates by conventional methods (29) and analysed by gas chromatography with flame ionization detector (GC-FID) and gas chromatography–mass spectrometry (GC-MS) as described above (See Chapter 2 of this thesis).

3.3.3 Bacteriophage assays

Propagation of bacteriophages on their respective host strains was performed as previously described (30). Similarly, both spot/plaque assays (31) and adsorption assays (32) were performed as previously described.

3.3.4 Cloning

All recombinant plasmids (Table 1) were generated in *L. lactis* NZ9000 and the primers (Table 2), unless otherwise indicated, were ordered from Eurofins MWG (Ebersberg, Germany). The two gene pairs of interest, *csdAB* and *csdEF*, were each amplified using the oligonucleotide pairs oIT25 and Oit28, and oIT37 and oIT40, respectively, with Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Massachusetts, USA) and cloned in the high copy number constitutive expression vector pNZ44 (18). The resulting constructs were named pNZ44::*csdAB* and pNZ44::*csdEF*. Due to the presence of pJP005 in *L. lactis* NZ9000-*cflA* that carries and chloramphenicol resistance marker, *csdAB* was also separately cloned into pNZ44str that carries a streptomycin resistance marker using the same oligonucleotides. Similarly, each of the genes from the two pairs was cloned independently into the pNZ44 vector, *csdA* using oligonucleotides oIT25-26, *csdB* using oIT27-oIT28, *csdE* using primers oIT39-oIT40, and *csdF* using primers oIT37-oIT38; the resulting constructs were named pNZ44::*csdA*, pNZ44::*csdB*, pNZ44::*csdE*, and pNZ44::*csdF*, respectively. Similarly each of the presumed promoter-containing regions located upstream of *csdAB*, (267 base pairs upstream of the annotated start codon of *csdA*) and *csdEF* (190 base pars upstream of the annotated start codon of *csdF*) was amplified using oligonucleotide pairs oIT29-oIT30 and oIT41-oIT42, respectively. The generated amplicons were individually cloned into the low copy number, promoter-probe vector pPTPL (19), resulting in pPTPL::*csdAB*-Prom and pPTPL::*csdEF*-Prom, where the second part of the plasmid name refers to the gene pair corresponding to the cloned promoter region.

Table 2. Oligonucleotides used in this study.

Oligo name	Sequence (5'-3')	Target/Comment
oIT19	ctattcccttactatatctctgcg	Fwd, 500 bp upstream of recombineering sequence site used for colony PCR check for knock-out mutants of *csdA*
oIT20	gtctgtccagtagttctttctctg	Rev, 500 bp downstream of recombineering sequence site used for colony PCR check for knock-out mutants of *csdA*
oIT21	cgataaaacttGAATTCG[1,3]	Fwd, oligonucleotide containing recombineering sequence used for colony PCR check for knock-out mutants of *csdA*
oIT22	tttattttcgtgggtaggttataagg	Fwd, 500 bp upstream recombineering sequence site used for colony PCR check for knock-out mutants of *csdB*
oIT23	gcttaatagagaaatgacaagacaaatc	Rev, 500 bp downstream of recombineering sequence site used for colony PCR check for knock-out mutants of *csdB*
oIT24	agatgctcaatGAATTCG	Rev, oligonucleotide containing recombineering sequence used for colony PCR check for knock-out mutants of *csdB*
oIT25	*aaaaaa*ccatggaatgagaataaaataatatg[2]	Fwd, 5'-end *csdA* gene for cloning in pNZ44
oIT26	*aaaaaa*tctagattagccttttgctcacaat	Rev, 3'-end *csdA* gene for cloning in pNZ44
oIT27	*aaaaaa*ccatggatagaaagaaggtttgattg	Fwd, 5'-end *csdB* gene for cloning in pNZ44
oIT28	*aaaaaa*tctagactacttagatttatccatga	Rev, 3'-end *csdB* gene for cloning in pNZ44
oIT29	*aaaaaa*ggatccggctttttttatttatattcaagac	Fwd, 267 bp upstream of 5'-end *csdA* for cloning in pPTPL
oIT30	*aaaaaa*tctagaattattttattctcattttcttag	Rev, 1 bp upstream of 5'-end *csdA* for cloning in pPTPL
oIT31	gacggaaattgtcctatac	Fwd, 500 bp upstream of recombineering sequence used for colony PCR check for knock-out mutants of *csdE*
oIT32	caatctcagatttgtagcc	Rev, 500 bp downstream of recombineering sequence used for colony PCR check for knock-out mutants of *csdE*
oIT33	tacgaacaatGATAACCC	Fwd, oligonucleotide containing recombineering used for colony PCR check for knock-out mutants of sequence *csdE*
oIT34	gcaactgtgttagcttttg	Fwd, 500 bp upstream of recombineering sequence used for colony PCR check for knock-out mutants of *csdF*
oIT35	gtaagaaaccaagtgcac	Rev, 500 bp downstream of recombineering sequence used for colony PCR check for knock-out mutants of *csdF*
oIT36	gcagcttattGATAACCC	Rev, oligonucleotide containing recombineering used for colony PCR check for knock-out mutants of sequence *csdF*
oIT37	*aaaaaa*ccatgggtataaaaaaagattctgtc	Fwd, 5'-end *csdF* gene for cloning in pNZ44
oIT38	*aaaaaa*tctagatcactccttaactctttttttc	Rev, 3'-end *csdF* gene for cloning in pNZ44
oIT39	*aaaaaa*ccatgggagttaaggagtgaatatg	Fwd, 5'-end *csdE* gene for cloning in pNZ44
oIT40	*aaaaaa*tctagattaatcttcgttaatgaagtc	Rev, 3'-end *csdE* gene for cloning in pNZ44
oIT41	*aaaaaa*ggatccgaaagacacaatcataaaaataaa	Fwd, 190 bp upstream of 5'-end *csdF* for cloning in pPTPL
oIT42	*aaaaaa*tctagataaagtaagacagaatctttttt	Rev, 1 bp upstream of 5'-end *csdF* for cloning in pPTPL

Recombineering		
oIT43	t*c*a*t*t*aatgtccggtcagtagaaccatcattgacaaaata aaCGAATTCaagttttatcgagcggtaacttttctgttttatc atctact[4]	*csdA*
oIT44	t*a*t*c*a*caaaactgattggggaagaatcaaccttagaaa accCGAATTCattgagcatctccactcatccagtcaacttttt gttagcgtaa	*csdB*
oIT45	a*a*c*a*t*aaacgattctacttttttggactaatgatatcattcg aaatGGGTTATCAttgttcgtatttggccttgagacgttcg gaagtttc	*csdE*
oIT46	g*c*a*a*a*aagtttttttgataatctaaaaatcccaataaaata GGGTTATCAataagctgctactaaaagaagatttcctaa cacaagtgca	*csdF*

Footnotes:[1]Capital letters: inserted sequence, [2]Italics: Non-genomic nucleotide sequence, [3]Underlined: Restriction site, [4]Asterisk: phosphorothioate modification

3.3.5 Recombineering

Recombineering was performed as previously described (17, 33, 34) with modifications to the assay adapted for use with *L. lactis* NZ9000 on the selected genes: *csdA*, *csdB*¸ *csdE*, and *csdF*. A total of 500 μg of mutational oligonucleotide (Table 2) inserting a stop codon and containing phosphorothioate linkages at the 5' end, was transformed into *L. lactis* NZ9000 pJP005. The recombineering oligonucleotides were produced by Integrated DNA Technologies (Leuven, Belgium).

3.3.6 Transcriptional analysis

The transcriptional pattern of *csdAB* and *csdEF* was investigated by measuring β-galactosidase activity produced by *L. lactis* NZ9000 pPTPL::*csdAB*-Prom and *L. lactis* NZ9000 pPTPL::*csdEF*-Prom during growth in M17 supplemented with 5 g/L glucose and 10 μg/mL tetracycline . The employed protocol is outlined in detail in Chapter 2 of this thesis.

3.3.7 Strain fitness, survivability, and biofilm formation

Viability/survival, fitness, and biofilm formation of strains derived from *L. lactis* NZ9000 carrying either mutations in *csdAB* or *csdEF*, or carrying high-copy number plasmids in which *csdAB* or *csdEF* had been cloned under the control of the strong, constitutive P44 promoter, were investigated following growth under various stress-inducing conditions. Initially, the

level of halotolerance (5.5 % w/v NaCl), ethanol tolerance (8 % v/v EtOH), tolerance to low pH conditions (pH 4.8), chelating agents (750 μM EDTA), and the bacteriocin nisin (15 μg/ml) was determined for *L. lactis* NZ9000 (Table 1) and based on these results the derived strains *L. lactis* NZ9000-*csdAB,* NZ9000-*csdEF,* NZ9000 pNZ44::*csdAB,* and NZ9000 pNZ44::*csdEF* were also tested. These strains were grown overnight and were then washed once with an equal volume quarter strength Ringer's solution. Subsequently, each of the strains was subcultured under the stress-inducing conditions listed above in a final volume of 200 μl with a 5 % inoculum. Growth was followed over a 24-48 hr period in a 96 well-plate. Biofilm formation was examined using a static microtiter plate assay as previously described (35). Briefly, the above-mentioned strains were grown overnight in M17 supplemented with 0.5 % w/v glucose (GM17) and were then diluted 1:100 in a 96 well-plate to 200 μl GM17; 200 μl of uninoculated GM17 was also included as a negative control. Following 24 hr incubation at 30 ^{o}C, biofilm formation was measured by a standard 0.05 % crystal violet staining protocol (36). The ability of each strain to form a biofilm was examined in triplicate.

3.4 Results

3.4.1 Identification of *csdAB* and *csdEF*

As described in Chapter 2 of this thesis, three gene pairs (*csdAB*, *csdCD*, and *csdEF*) had been identified based on the similarly of their deduced amino acid sequences and associated membrane topology to proteins involved in LTA and WTA decoration and characterised in *Listeria monocytogenes* (37, 38). Mutants of *csdCD* were shown to exhibit structural differences in the PSP component of their CWPS, i.e., the enzymes encoded by *csdCD* were shown to be responsible for Glc attachment of the side chain Glc of the PSP (Chapter 2). In order to assess if *csdAB* and *csdEF* are involved in glycosidic modifications of other cell envelope component(s), derivatives of *L. lactis* NZ9000 and 3107 were constructed that harboured either knock-out mutations in or that contained vectors (over)expressing the two gene pairs together or independently (*csdAB* or *csdEF*) (Table 1). These mutants were then subjected to analyses focusing on altered cell envelope-associated glycan structures and any apparent phenotypic effects (as outlined below).

3.4.2 CWPS structure analysis of *L. lactis* NZ9000 *csdAB* mutants and overexpressing strains

CWPS was extracted from bacterial cells by HF treatment and purified by SEC-HPLC allowing separation of rhamnan chains and PSP oligosaccharide subunits as described previously (9). The monosaccharide composition of the CWPS of the mutant strains was analysed by ion-exchange chromatography and their structures by Maldi-TOF MS.

Although low amount of Glc was detected in the rhamnan component of wild-type *L. lactis* NZ9000 (Glc/Rha ~ 1/27), a decrease of this Glc/Rha ratio (from ~1/27 to ~ 1/40) was observed in the rhamnan component purified from *L. lactis* NZ9000-*csdAB* (Table 3) while no visible effect could be observed in the complex mass spectrum of the rhamnan from such

mutants. Furthermore, the PSP composition was unchanged in *L. lactis* NZ9000-*csdAB* compared to WT.

In line with the results obtained for *L. lactis* NZ9000-*csdAB*, we demonstrated that, with *csdAB* overexpression in *L. lactis* NZ9000, a huge increase of Glc/Rha ratio (from ~ 1/27to ~1/3; Table 3) was detected in the rhamnan component. Maldi-TOF MS analysis confirmed a mass increase of the major rhamnan chain (Figure 1A). In *L. lactis* NZ9000 pNZ44, the major signal (m/z 5061.8) was tentatively assigned to $[M+Na]^+$ adducts of HexNAc-Rha_{33} as described previously (9). In the spectrum of the overexpressing strain, the signal at m/z 6356.5 could be assigned to $[M+Na]^+$ adducts of HexNAc-Rha_{33}-Glc_8 and presenting a mass increase of 1296-mass unit (m.u.) corresponding to eight additional Glc residues compared to the major peak (m/z 5061.8) of the control strain. Also other signals could be attributed to $[M+Na]^+$ adducts of rhamnan chains with an increased amount of Glc substituents (Figure 1A). Methylation analysis confirmed a large increase of terminal Glc (t-Glc) in the rhamnan purified from *L. lactis* NZ9000 pNZ44::*csdAB*, while a 2,3-disubstituted Rha appeared and the levels of 3-Rha markedly decreased compared to the control NZ9000 rhamnan. These results confirmed the presence of Glc substitutions on rhamnan and identify the site of branching. From the methylation analysis, we can tentatively deduce the following structure for the subunits substituted with Glc: {→2)-α-Rha-(1→2)-α-Rha-(1→3)-[Glc-(1→2)]-α-Rha-}. No effect of *csdAB* overexpression was detected on PSP structure either by composition or MS analysis (Figure 3).

Overexpression of *csdA* alone in NZ9000 did not modify the rhamnan or PSP structures. In contrast, overexpression of *csdB* alone in NZ9000 was shown to result in a huge increase of Glc in the rhamnan component with a ratio Glc/Rha reaching ~1/4 (Table 3) and a MS spectrum showing higher molecular mass species (Figure 4) similar to *csdAB* overexpression. Interestingly, the MS spectrum of PSP from NZ9000 overexpressing only *csdB* also appeared

modified, with an increased intensity of the peak corresponding to pentasaccharide versus the hexasaccharide peak (Figure 3). This result was corroborated by methylation analysis that showed a decrease of t-Glc relative to Glc in the PSP, thus indicating a reduction of Glc side chain addition in PSP in this latter strain (Figure 1). These results suggest that part of the C_{55}-P-Glc intermediate synthesized by CsdC, endowed with the same C_{55}-P-Glc synthase activity as CsdA (see Chapter 2), yet involved in PSP glucosylation is titrated away by CsdB when this protein overexpressed, and directed to perform rhamnan glucosylation.

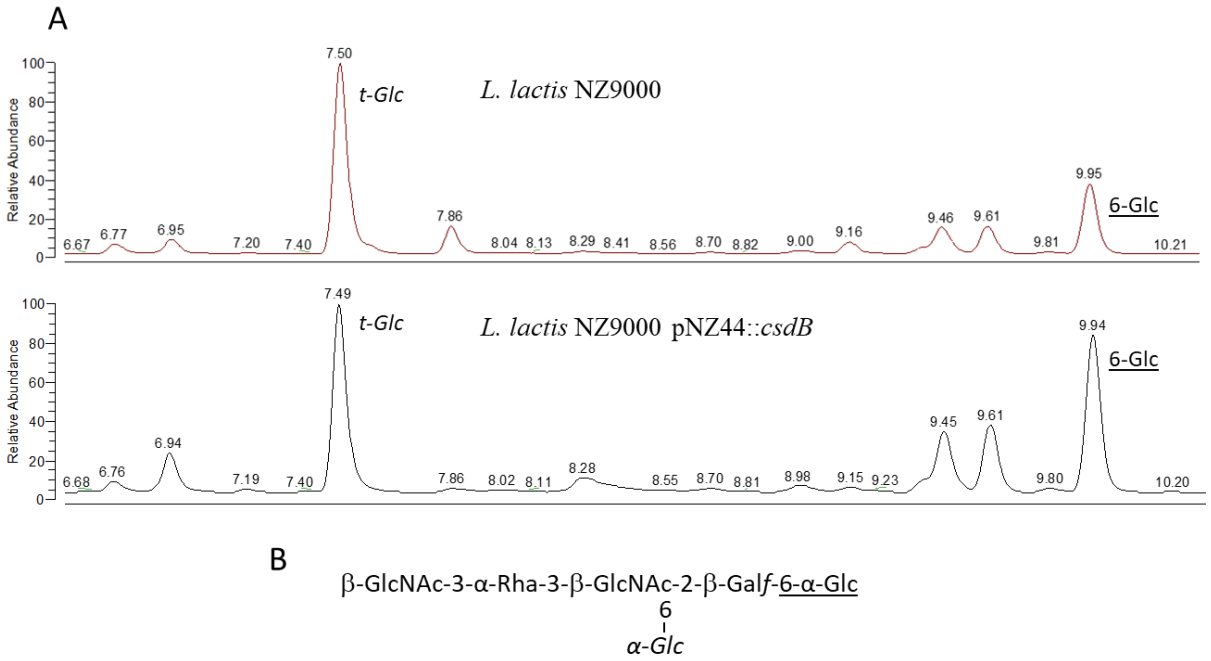


Figure 1. [A] Methylation analysis of PSP extracted from *L. lactis* NZ9000 and NZ9000 pNZ44::*csdB*. Samples were analysed by GC-MS. [B] Figure highlighting a single subunit of the *L. lactis* PSP structure. Peaks identified by *italics* or underlined fonts in [A] or also highlighted in this schematic.
t-Glc = Terminal glucose

3.4.3 Expression of *csdAB* genes in *L. lactis* 3107 and CWPS structure analysis

Since the *csdAB* gene pair appears to be absent in the genome of *L. lactis* 3107, the *L. lactis* NZ9000-derived *csdAB* genes were cloned and expressed in strain 3107 in order to evaluate their role and impact on this strain. Compositional analysis of the PSP of *L. lactis* 3107 pNZ44::*csdAB* appeared identical to the control (*L. lactis* 3107 pNZ44) and wild-type *L. lactis* 3107. In contrast, the composition of the purified rhamnan component of the CWPS was shown to exhibit a very substantial increase of Glc/Rha ratio reaching a value close to 1/2, whereas only traces of Glc were found in the rhamnan peak of control *L. lactis* 3107 containing the empty plasmid pNZ44 (Table 3). Maldi-TOF MS analysis confirmed a clear mass increase for the rhamnan extracted from *L. lactis* 3107 pNZ44::*csdAB* compared to the control *L. lactis* 3107 pNZ44 and wild-type strains (Figure 1B). The two major peaks of the spectrum of the wild-type rhamnan with m/z values of 4723.0 and 5162.0 were tentatively assigned to $[M+Na]^+$ adducts of Rha_{32} and Rha_{35} with a 438-m.u. difference corresponding to the subunit ([-2-α-Rha-2-α-Rha-3-α-Rha-]) described previously (9). In the *L. lactis* 3107 pNZ44::*csdAB* strain, the two major peaks at m/z values of 6021.0 and 6622.4 presents a 601-m.u. difference corresponding to 3 Rha plus 1 Glc. These two signals could be assigned to $[M+Na]^+$ adducts of Hex_9-$(Rha)_{31}$ and Hex_{10}-$(Rha)_{34}$, presenting a mass difference of 1296-m.u. and 1458-m.u., respectively with the two major peaks of the wild-type 3107 rhamnan spectrum, corresponding to the addition of 8 and 9 Glc residues, respectively. When *csdA* or *csdB* was expressed individually in *L. lactis* 3107, no difference in the rhamnan composition or MS spectrum was observed (Table 3), indicating that both *csdA* and *csdB* are required for rhamnan glucosylation in 3107. Together, these results imply that the two proteins CsdA and CsdB are functional when heterologously expressed in *L. lactis* 3107 and are able to add Glc substituents on rhamnan.

3.4.4 CWPS structure of *L. lactis cflA* negative mutants

As described in chapter 2, we have identified by sequence homology search on the *L. lactis* NZ9000 genome, a gene (*llnz_02975* renamed *cflA*) encoding a putative flippase with 26 % amino acid sequence identity with the *Ls. monocytogenes* GtcA flippase (38) and endowed with four putative TMHs. Such a flippase is proposed to be involved in the transfer of the C55-P-Glc intermediate from the inner to the outer side of the cytoplasmic membrane as part of a three-component glycosylation system (39). We have shown in chapter 2, that CflA is involved in Glc side-chain addition on PSP subunits. Since no other flippase candidate could be identified by BLAST homology search in the NZ9000 genome, we also investigated the role of CflA in Glc addition on rhamnan. The *cflA* gene was knocked out in *L. lactis* NZ9000 as well as in strain NZ9000 pNZ44str::*csdAB* synthesizing rhamnan with a high level of Glc substituents (see above). CWPS analysis was performed after extraction by HF and SEC-HPLC separation of rhamnan and PSP oligosaccharides as described in Material and Methods.

Inactivation of *cflA* in wild-type NZ9000 led to a decreased amount of Glc in the purified rhamnan peak (Table 3), suggesting the involvement of *cflA* in Glc substituent addition onto rhamnan. To investigate further the role of *cflA*, the *csdAB* genes were overexpressed in NZ9000-*cflA* mutant. Whereas, as described above, overexpression of *csdAB* in wild-type NZ9000 background led to a very high level of Glc grafted to rhamnan (Glc/Rha ~3, Table 3), overexpression of *csdAB* in NZ9000-*cflA* had no (or very low) impact on the Glc substitution level of rhamnan (Table *3*). These results indicate that CflA is required for glucosylation of rhamnan, and could act as the flippase necessary for the transfer of C_{55}-P-Glc intermediate synthesized by CsdA, from the inner to the outer side of the cytoplasmic membrane, to be the substrate of CsdB for adding Glc onto rhamnan.

In conclusion, our results identify CsdA, CsdB and CflA as a three-component glucosylation system of rhamnan in *L. lactis*. Moreover, according to our results described in

chapter 2, CflA can also work in concert with CsdC and CsdD to add side chain Glc onto the PSP subunit, although it seems partially dispensable for this process, thus suggesting the involvement of an alternative protein with flippase activity in *L. lactis* NZ9000.

Table 3. Monosaccharide composition of the rhamnan peak purified by HP-SEC after HF extraction from the different *L. lactis* strains. Values are standardized on Glc

Strain	Rha[1]	Glc	GlcNAc[2]	Gal[2]
L. lactis NZ9000	1	0.059	0.012	0.035
L. lactis NZ9000-*csdAB*	1	0.032	0.013	0.039
L. lactis NZ9000-*csdA*	1	0.025	0.010	0.035
L. lactis NZ9000-*csdB*	1	0.036	0.007	0.036
L. lactis NZ9000 pNZ44	1	0.067	0.007	0.013
L. lactis NZ9000 pNZ44*::csdAB*	1	0.333	ND[3]	ND
L. lactis NZ9000 pNZ44*::csdA*	1	0.040	0.008	0.012
L. lactis NZ9000 pNZ44*::csdB*	1	0.250	0.025	ND
L. lactis 3107 pNZ44	1	0.038	0.050	0.031
L. lactis 3107 pNZ44*::csdAB*	1	0.500	0.050	0.050
L. lactis 3107 pNZ44*::csdA*	1	0.042	0.050	0.038
L. lactis 3107 pNZ44*::csdB*	1	0.042	0.042	0.025
L. lactis NZ9000-*cflA* pNZ44str	1	0.027	0.056	0.056
L. lactis NZ9000 pNZ44str*::csdAB*	1	0.250	0.100	0.050
L. lactis NZ9000-*cflA* pNZ44str*::csdAB*	1	0.031	0.063	0.053

Footnotes:[1]Values are standardized relative to Rha, [2]GlcNAc and Gal are found in the rhamnan peak but thought to be PSP contaminants, [3]ND – not detected

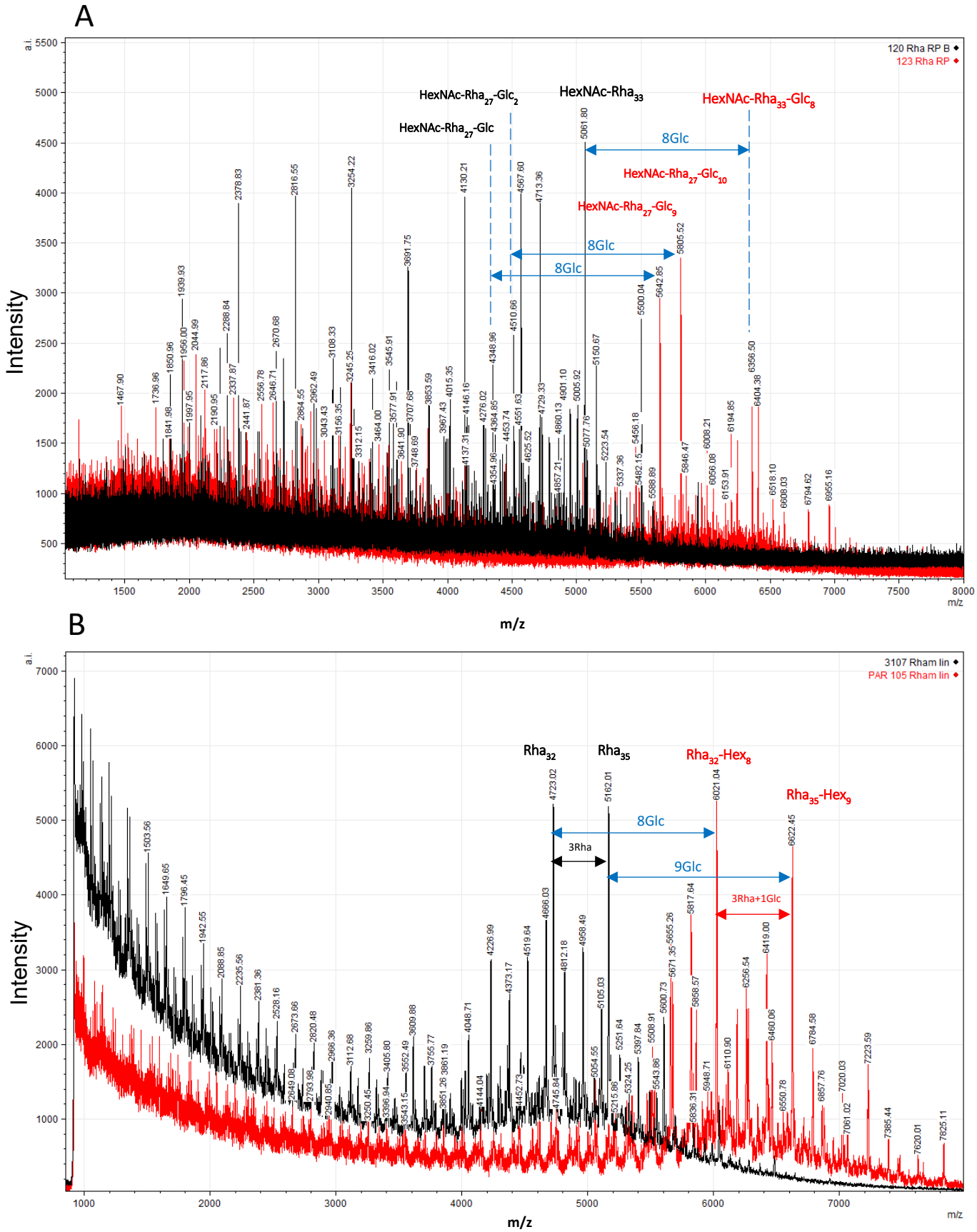


Figure 1. Maldi-TOF MS spectra of rhamnan purified from: (A) *L. lactis* NZ9000 pNZ44 (black) and *L. lactis* NZ9000 pNZ44::*csdAB* (red); (B) *L. lactis* 3107 Pnz44 (black) and *L. lactis* 3107 pNZ44::*csdAB* (red); m/z values correspond to $[M+Na]^+$ adducts.

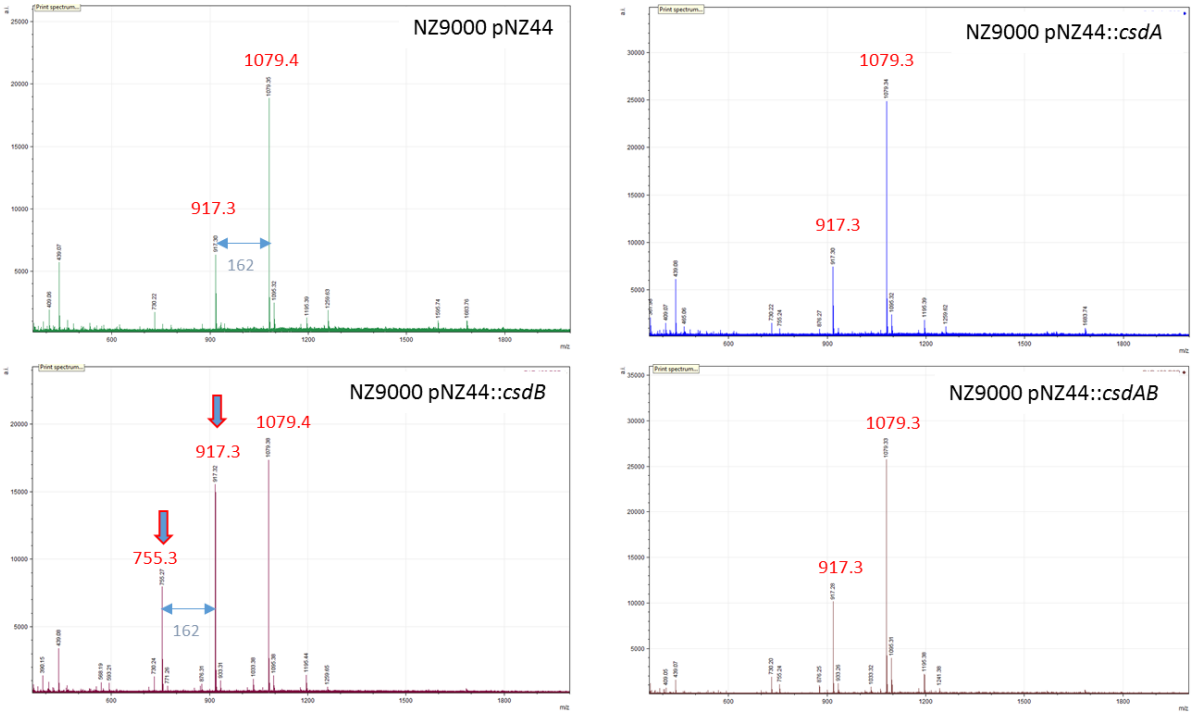


Figure 3. Maldi-TOF MS spectra of PSP oligosaccharides purified by HP-SEC after HF extraction from *L. lactis* NZ9000 pNZ44 and the different mutant strains overexpressing *csdA*, *csdB* or both genes *csdA* and *csdB*. m/z values correspond to $[M+Na]^+$ adducts.

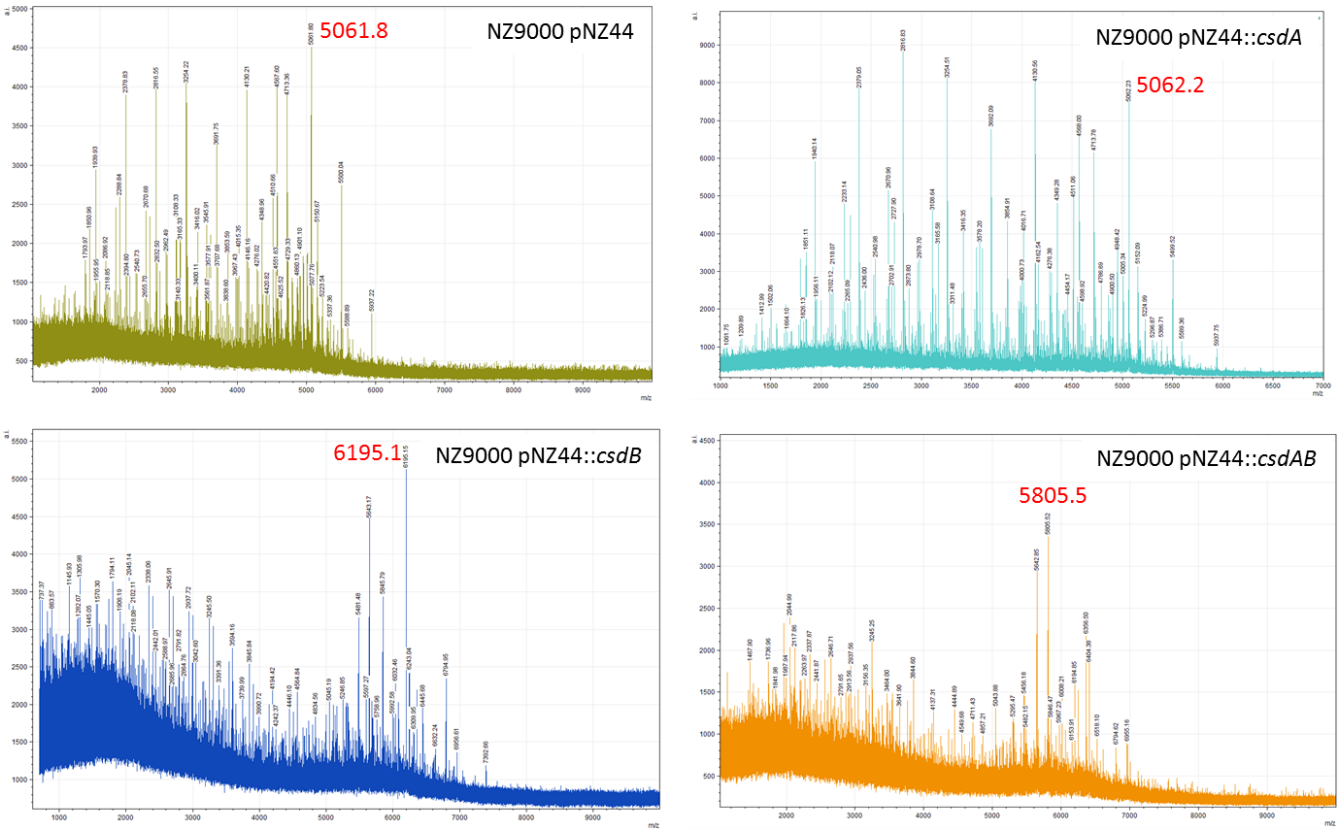


Figure 4. Maldi-TOF MS spectra of rhamnan purified from control *L. lactis* NZ9000 pNZ44 and mutant *L. lactis* NZ9000 overexpressing strains NZ9000 pNZ44::*csdA*, NZ9000 pNZ44::*csdB* and NZ9000 pNZ44::*csdAB*. In red, m/z value corresponding to the peak with the highest intensity for each spectrum. m/z values correspond to $[M+Na]^+$ adducts.

3.4.5 CWPS structure analysis of *L. lactis* NZ9000-*csdEF* and overexpressing strains

In light of the identified functions of CsdAB and CsdCD in the sugar decoration of the lactococcal rhamnan and PSP structures, we set out to identify the function of the third closely related gene pair *csdEF*. No modification of the CWPS components (rhamnan or PSP) was detected in *L. lactis* NZ9000-*csdEF* or in *L. lactis* NZ9000 pNZ44::*csdEF* (data not shown). These results suggest that CsdE and CsdF could be involved in the addition of a monosaccharide onto another cell wall glycopolymer such as WTA or LTA. Of note, *L. lactis* was previously shown to synthesize LTA that consists of glycerol-phosphate chains, with Gal as sugar substituent (40). In order to verify this hypothesis, LTA was extracted from the WT strain *L. lactis* NZ9000 and its mutant NZ9000-*csdEF* with hot 80 % phenol (26). Of note, the use of phenol for the isolation of LTAs should guarantee a relative purity of the sample, however, the possibility of cross-contamination from other TA structures, such as WTA, cannot be excluded. The degree of Gal glycosylation of LTA was assessed based on relative quantities of glycerol, Gal and terminal Gal in two preparations. We have found for the same amount of glycerol, the amount of Gal decreased at least ten times in the *L. lactis* NZ9000-*csdEF* mutant preparation (Figure 5). In agreement with these results, the amount of terminal Gal decreased dramatically in this sample, compared to the preparation from the WT strain *L. lactis* NZ9000 (Figure 6). All these data clearly indicate that inhibiting the expression of both *csdE* and *csdF* affects the galactosylation of poly(glycerophosphate) chains of LTA in *L. lactis* NZ9000.

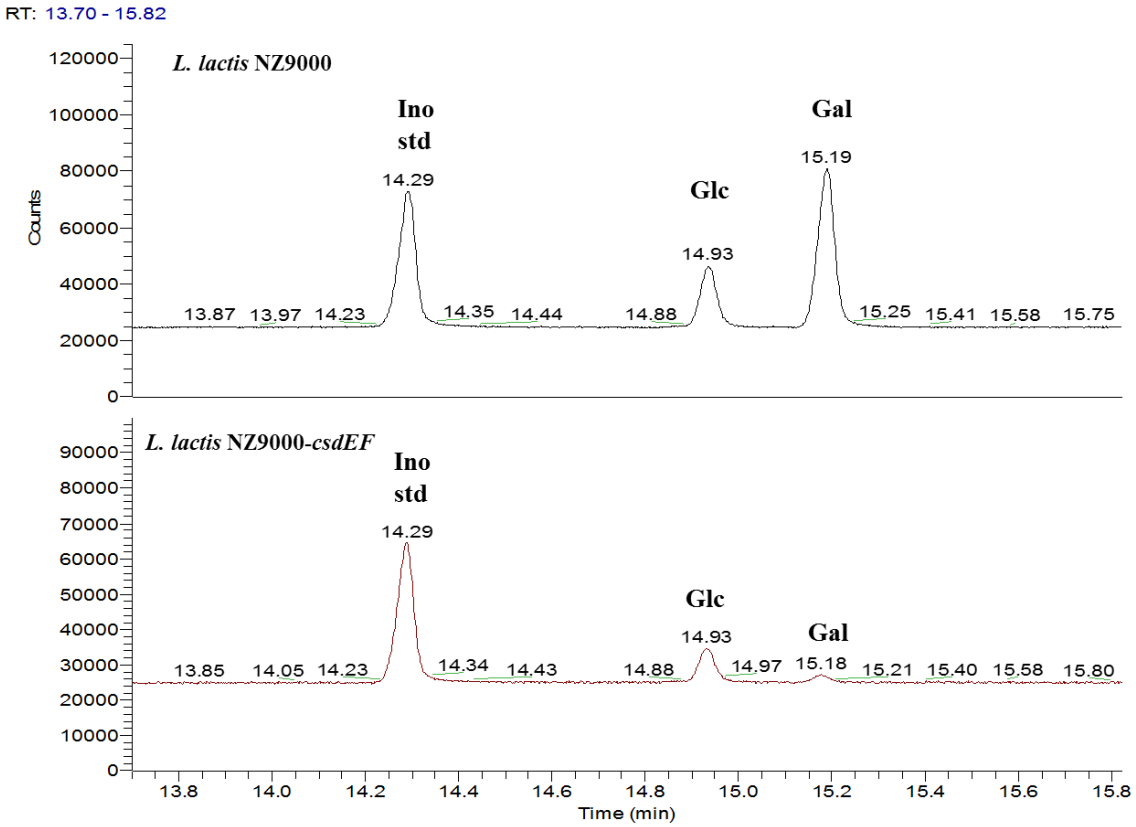


Figure 5. Monosaccharide analysis of crude phenol extracts from *L. lactis* NZ9000 and NZ9000-*csdEF*. 5 mg of each extract, along with Inositol (Ino, 90 µg) as an internal standard, were treated with HF, hydrolysed, converted into alditol acetates and analysed by GC-FID. The figure has been zoomed onto the region corresponding to hexoses.

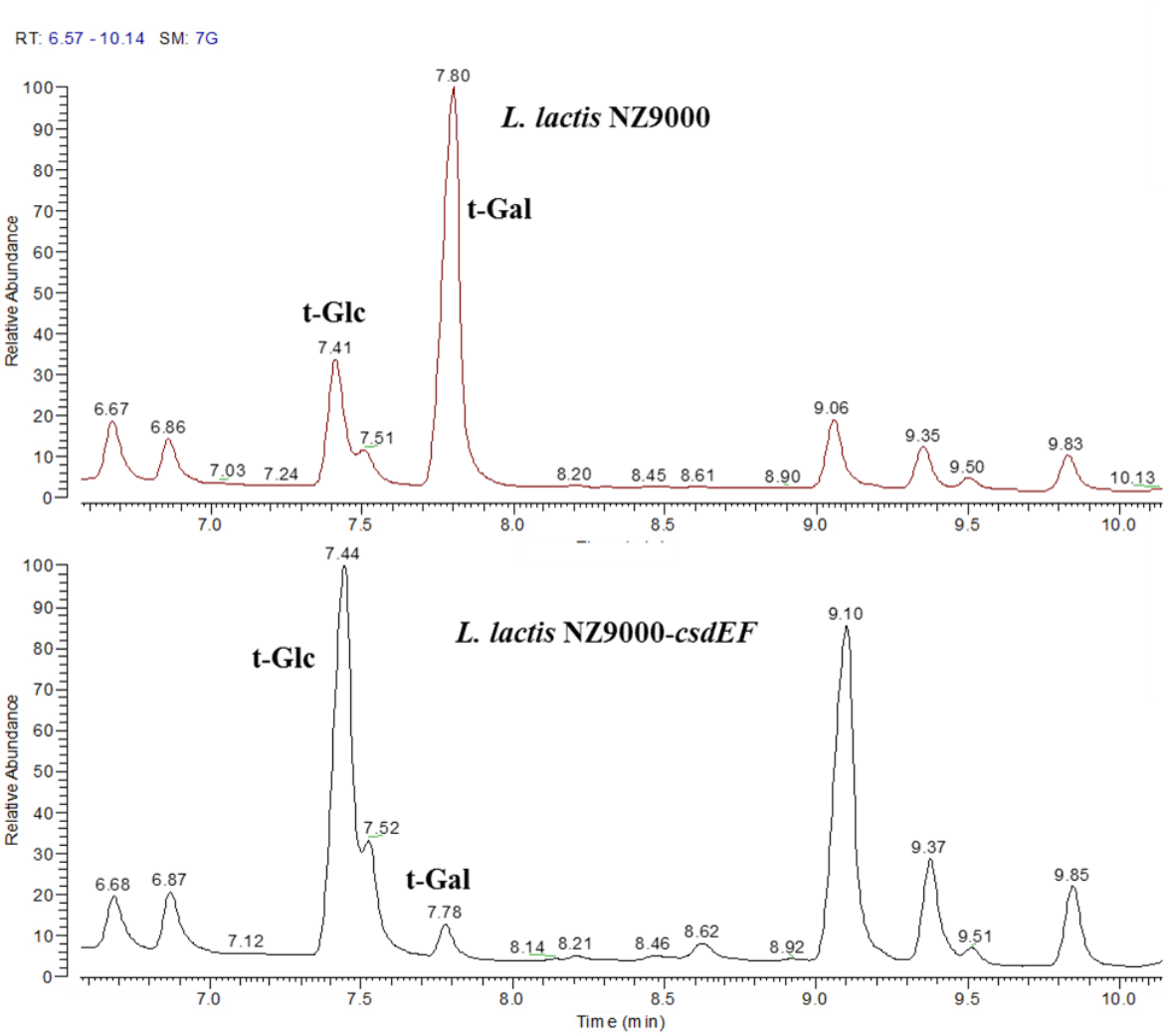


Figure 6. Methylation analysis of crude phenol extracts from *L. lactis* NZ9000 and NZ9000-*csdEF*. Samples were analysed by GC-MS. t-Glc = Terminal glucose; t-Gal = Terminal galactose.

3.4.6 Role of *csdAB* and *csdEF* in bacteriophage infection

In order to determine the extent to which the above-mentioned structural modifications of the lactococcal CWPS, mediated by the altered expression of *csdAB* and *csdEF*, were relevant to any particular phenotype, we examined the interactions of all the derived strains against a set of strain-specific phages (Table 1). None of the *L. lactis* NZ9000-derived strains, i.e. *L. lactis* NZ9000-*csdAB*, NZ9000-*csdEF*, NZ9000 pNZ44::*csdAB*, and NZ9000 pNZ44::*csdEF*, exhibited altered phage sensitivity profiles (four 936-type phages tested, Table 1). Similarly, overexpression of *csdEF* in *L. lactis* 3107 did not affect phage-host interactions in any significant way (six 936-type and seven P335-type phages tested, Table 1). Overexpression of

csdAB in *L. lactis* 3107 caused reduced sensitivities against both P335 and 936 type phages (Figure 7). In particular, expression of *csdAB* was shown to cause complete resistance to phage 07501. Furthermore, a decreased plaque size and altered plaque appearance was observed for phages LC3 (P335), 07501 (P335), 66902 (936), 66903 (936), which were all shown to exhibit reduced infectivity against *L. lactis* 3107 pNZ44::*csdAB* (Table 4). Similar results were observed when *csdAB* was overexpressed in a *L. lactis* VES5751, a PSP-deficient derivative of *L. lactis* MG1363 that is resistant to infection by most phages commonly infecting the wild-type strain (6). Two phages, jj50* and MCC1, derivatives of phage jj50 (Genbank accession no. NC_00837) and sk1 (Genbank accession no. NC_001835.1) respectively, are unaffected by the PSP-deficient phenotype of *L. lactis* VES5751 and can maintain their ability to infect it. However, upon overexpression of *csdAB* in *L. lactis* VES5751 the infectivity of jj50* is reduced by 2 logs compared to the control strains (Figure 8), while also producing smaller plaques size and altered plaque appearance (Table 5). Phage MCC1 appears to be unaffected by *csdAB* overexpression (Figure 8), however, a slight decrease in plaque size is observed (Table 5)

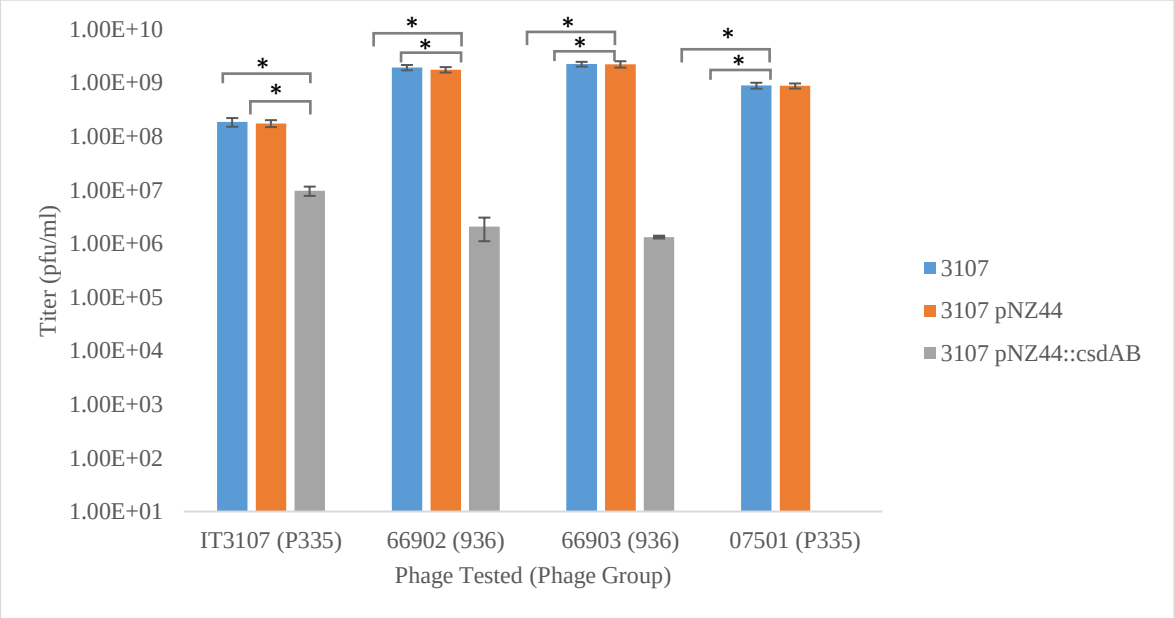


Figure 7. Total infectivity of phages LC3, 66902, 66903, and 700 against *L. lactis* 3107, 3107 pNZ44, and 3107pNZ44::*csdAB*. p values are indicated by stars *p<0.001

Table 4. Plaque size and morphology of phages LC3, 66902, 66903, and 700 against strains *L. lactis* 3107 and *L. lactis* 3107 pNZ44::*csdAB*

		***L. lactis* 3107**	***L. lactis* 3107 pNZ44::*csdAB***
Phage Tested	LC3	1-1.5 mm (fuzzy halo)	0.5 mm
	66902	2.5-3 mm (no halo)	<0.5 mm (difficult to discern)
	66903	2.5-3 mm (no halo)	<0.5 mm (difficult to discern)
	07501	<1 mm (with fuzzy halo)	(-)

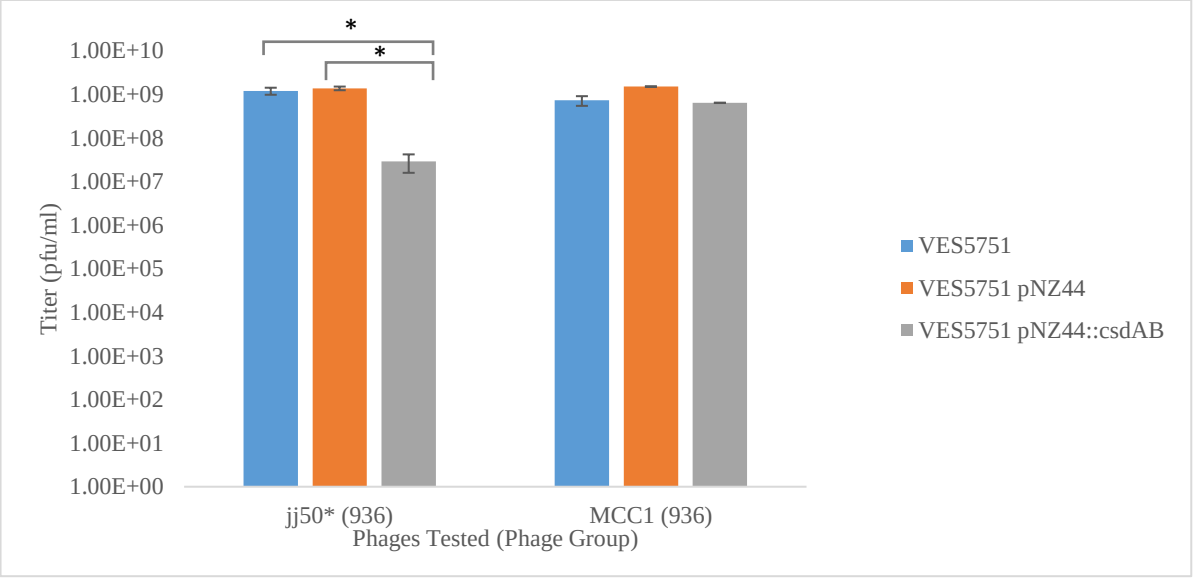


Figure 8. Total infectivity of phages jj50* and MCC1 against *L. lactis* VES5751, VES5751 pNZ44, and VES5751 pNZ44::*csdAB*. p values are indicated by stars *p<0.001.

Table 5 Plaque size and morphology of phages jj50* and MCC1 against strains *L. lactis* VES5751 and *L. lactis* VES5751 pNZ44::*csdAB*

		***L. lactis* VES5751**	***L. lactis* VES5751 pNZ44::*csdAB***
Phages Tested	jj50*	1.5-2 mm (halo)	<0.5 mm (very faint plaques)
	MCC1	1.5-2 mm (halo)	~1 mm (halo)

3.4.7 Effect of modulating expression of *csdAB* on bacteriophage adsorption

Following the finding that overexpression of *csdAB* causes altered phage sensitivities, we investigated if this phenotype may be due to interference with the phage's ability to adsorb to its host. Despite the apparent effect of *csdAB* overexpression in *L. lactis* 3107 on phage

infectivity, the effect appeared to be unrelated to phage adsorption as the rate of adsorption of both the P335 (LC3 and 07501) and 936 (66902 and 669023) type phages was similar on both wild-type and derivative strains of *L. lactis* 3107 (data not shown). In contrast, overexpression of *csdAB* and associated reduced phage infectivity in *L. lactis* VES5751 does appear to be associated with decreased phage adsorption levels, as illustrated in Figure 9.

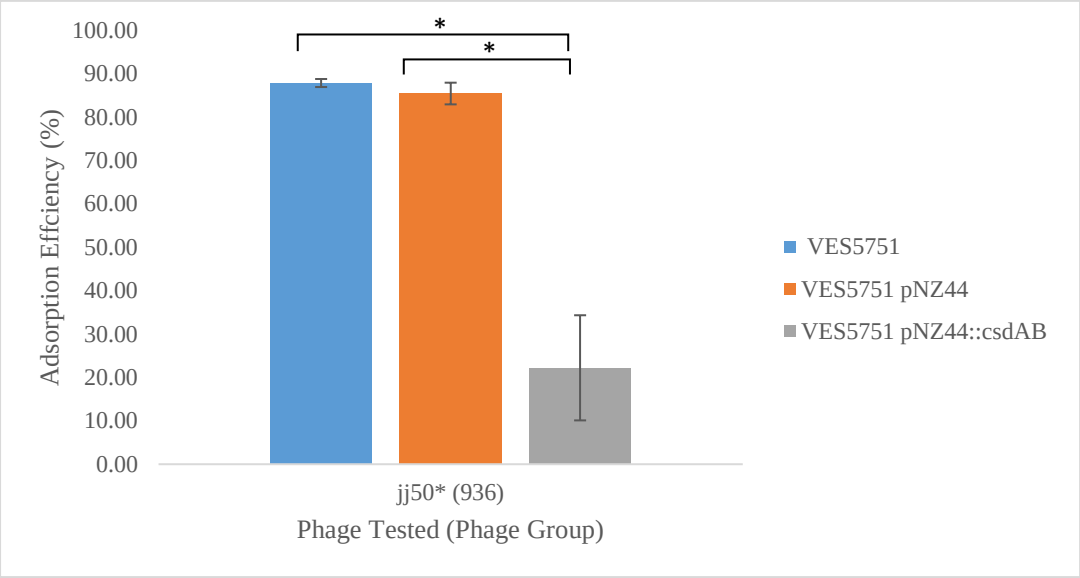


Figure 9. Percent adsorption efficiency of phages jj50* against *L. lactis* VES5751, VES5751 pNZ44, and VES5751 pNZ44::*csdAB*. p values are indicated by stars *p<0.001.

3.4.8 Transcriptional analysis of *csdAB* and *csdEF*

To assess the transcriptional activity of the predicted promoters of *csdAB* and *csdEF*, the complete upstream intergenic sequence of the genes was cloned into the promoter probe vector pPTPL (Table 1). The two derived plasmids pPTPL::*csdAB*-Prom and pPTPL::*csdEF*-Prom were individually transformed into *L. lactis* NZ9000 and the level of β-galactosidase activity was monitored across a time period of 6.75 to 8.5 hours. The β-galactosidase activity of the *csdAB*-Prom-*lacZ* fusion does not exceed 38.2 Miller units during the exponential growth phase of the strain (Figure 10) while the *csdEF* promoter activity appears to peak during the lag phase

immediately following sub-cultivation, with an average of 1333.2 Miller units, and steadily decreases across the ensuing exponential and stationary phases (Figure 11).

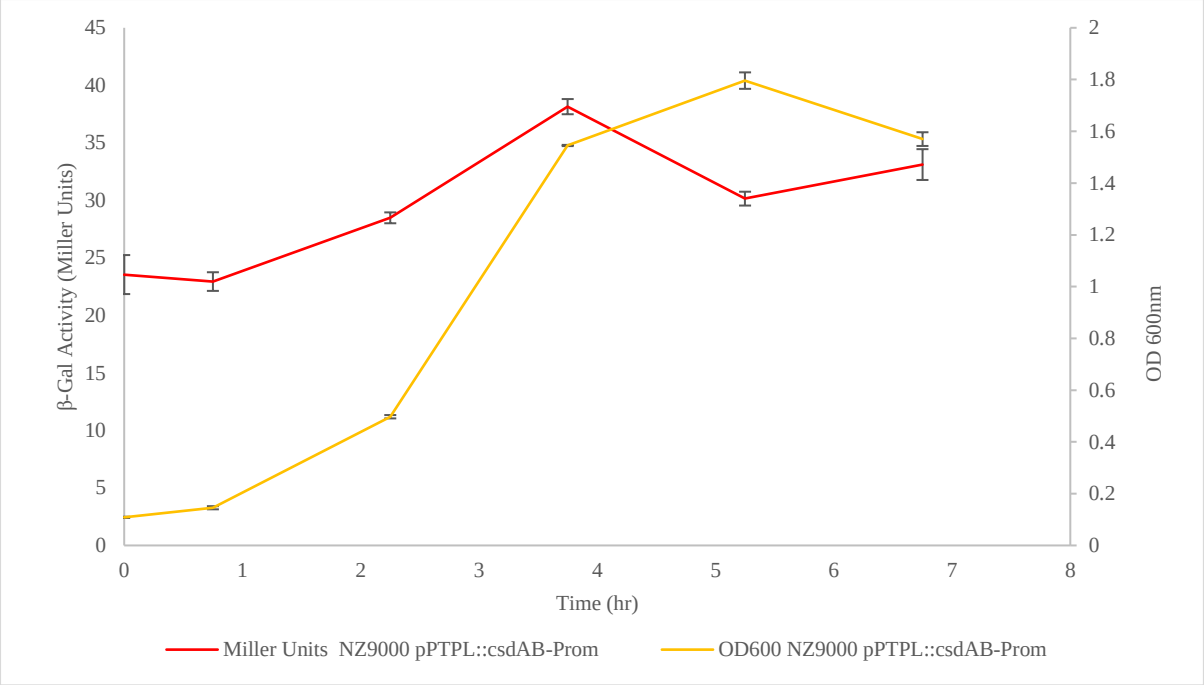


Figure 10. Specific β-Galactosidase production during normal growth of the *L. lactis* NZ9000 strain carrying a promoter-less pPTPL vector and a pPTPL vector carrying the *csdAB* upstream intergenic region (*csdAB*-Prom).

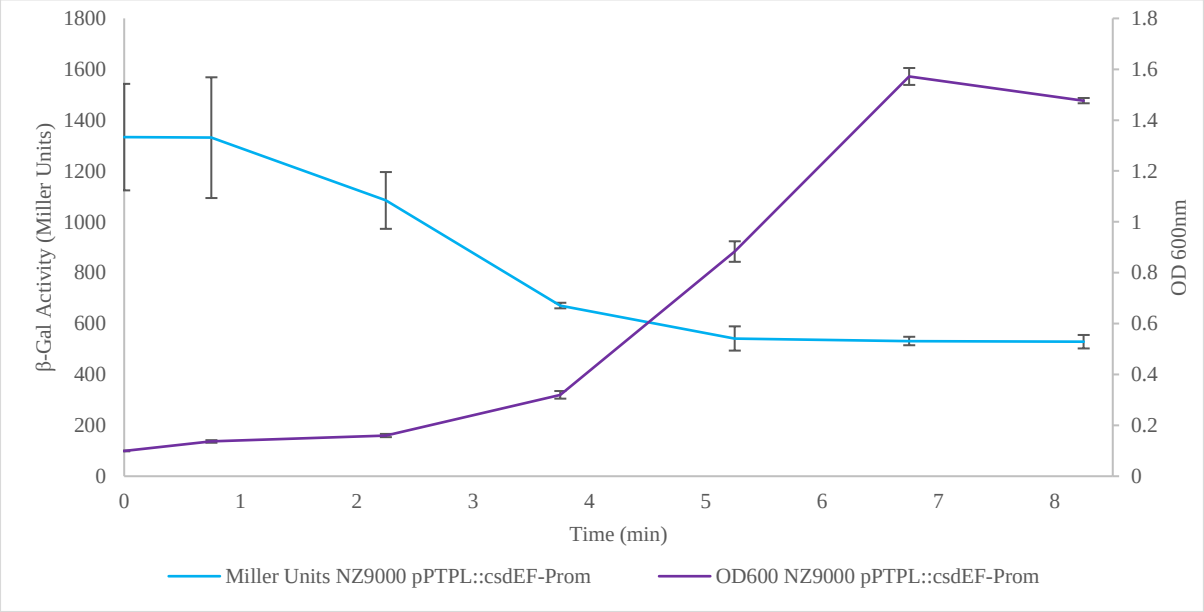


Figure 11 Specific β-Galactosidase production during normal growth of the *L. lactis* NZ9000 strain carrying a promoter-less pPTPL vector and a pPTPL vector carrying the *csdEF* upstream intergenic region (*csdEF*-Prom).

3.4.9 Strain fitness and biofilm formation with *csdAB* and *csdEF* modulated expression

Additional phenotypic effects due to the mutation or overexpression of *csdAB* and *csdEF* were investigated. All derived *L. lactis* NZ9000 strains were grown under a number of stress-inducing conditions by incorporating growth-limiting concentrations of NaCl, ethanol, EDTA, and reduced pH. It was observed that under normal growth conditions (M17 + 0.5 % w/v Glu) knock-out mutations and overexpression of these two sets of genes did not lead to any detrimental effect on growth rate. The stressors, with the exception of the antimicrobial peptide nisin, appeared to similarly affect both the wild-type and the derivative strains. Any deviations in the growth pattern of the wild-type and derivative strains were demonstrated to be statistically insignificant (data not shown). In the presence of nisin in the growth medium, the knock-out mutation and overexpression of *csdEF* caused altered growth patterns. For example, overexpression of the *csdEF* appears to provide increased resistance to the antimicrobial, while the opposite effect is observed in the knock-out mutant (Figure 12). Overexpression of the two gene pairs, *csdAB* and *csdEF*, in *L. lactis* NZ9000 also leads to divergent biofilm formation phenotypes (Figure 13). Overexpression of *csdAB* significantly decreases the ability to form biofilm in relation to the control strain, while overexpression of *csdEF* has the opposite effect on the strain's biofilm forming abilities.

GM17
+15 µg/ml Nisin

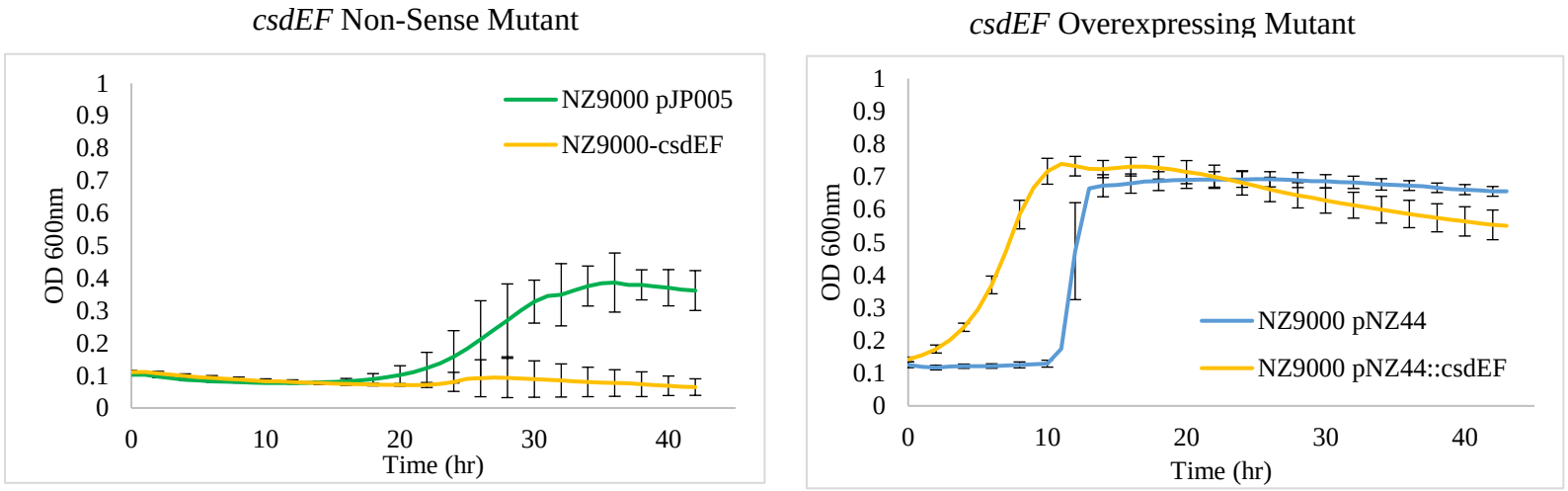


Figure 2. Kinetic growth *L. lactis* NZ9000 controls and *csdEF* derivatives under stress inducing condition.

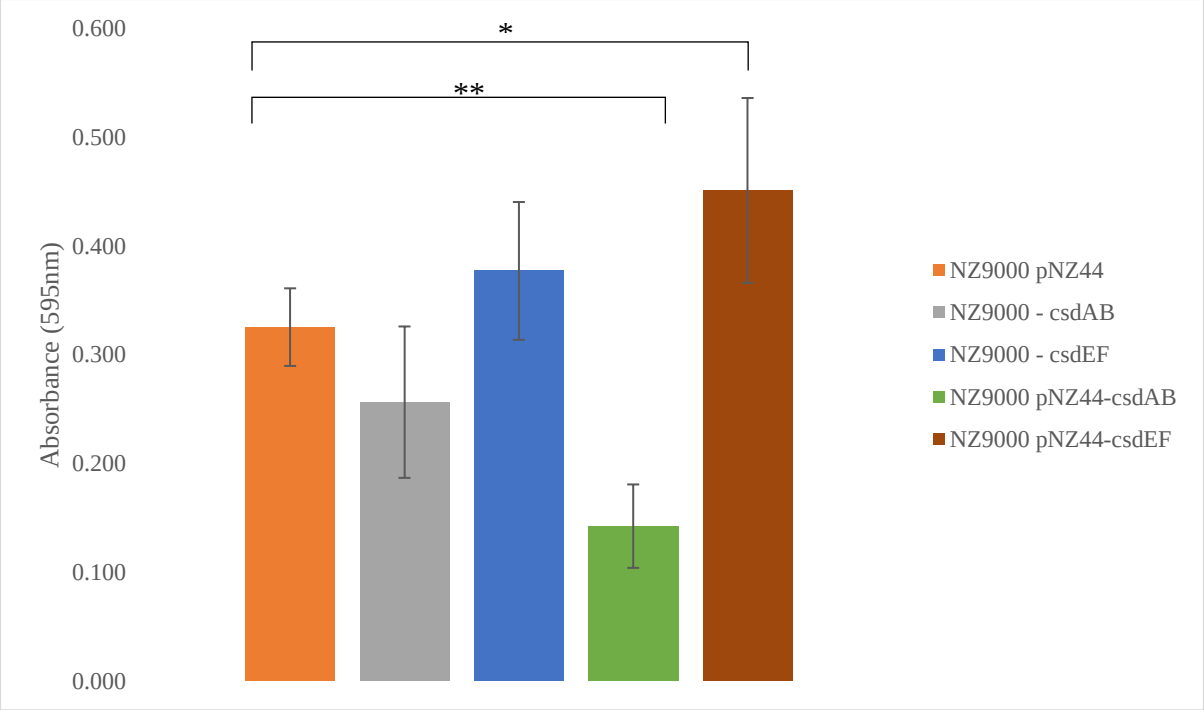


Figure 13. Biofilm formation of *L. lactis* NZ9000 and its derivatives following 36 hr incubation. p values are indicated by stars *p<0.01 and **p<0.0001

3.5 Discussion

The link between the *L. lactis* CWPS genotype and the structural diversity of the corresponding CWPS structures has previously been described (6-8). Similarly, we have recently highlighted the existence of a pair of adjacent genes along with a flippase-encoding gene, which are unconnected to the *cwps* gene cluster. This gene pair is involved in the structural modification of the CWPS in *L. lactis* NZ9000 (Chapter 2). These genes, *csdC* and *csdD* (together designated *csdCD*) and the flippase-encoding gene *cflA*, were shown to be involved in the glucosylation of the PSP structure of *L. lactis* and this modification was tentatively linked to phage resistance/sensitivity. Based on exhibited homologies, two additional gene pairs were identified within the *L. lactis* NZ9000 genome and were investigated in this chapter.

Here, we show that the products of *csdAB* add a glucosyl moiety to the rhamnan subunit and that the enzymatic acitivities of both proteins are required for this addition (Figure 14). According to methylation analysis, Glc is added specifically on 3-Rha of the [-2-α-Rha-2-α-Rha-3-α-Rha-] subunit. A chromosomal copy of *csdAB* is present in the genome of *L. lactis* NZ9000 while it is absent in *L. lactis* 3107. Overexpression of *csdB*, which encodes a polytopic membrane glycosyltransferase (PolM GT) in NZ9000, also allowed grafting of one Glc side chain on almost every 3-Rha subunit as was observed upon overexpression of *csdAB*. These results suggest that CsdB can utilise the undecaprenyl-phosphate-glucose (C_{55}-P-Glc) substrate produced by another C_{55}-P-Glc synthase apart from that produced by its cognate partner CsdA. A possible candidate is CsdC, which is another C_{55}-P-sugar synthase involved in glucosylation of PSP (see Chapter 2) in collaboration with CsdD,. These observations confirm that CsdB provides specificity of the glucosylation reaction towards rhamnan. Interestingly, we have shown previously (see Chapter 2) that in the absence of CsdC (C_{55}-P-Glc synthase), CsdD (PolM GT) is still able to glucosylate PSP, also suggesting the use by CsdD of C_{55}-P-Glc

substrate produced by another C_{55}-P-Glc synthase, the obvious candidate for this being CsdA. Additionally, the gene pair *csdEF* has been implicated in the galactosylation of the LTA structures of *L. lactis* NZ9000. Introduction of non-sense mutations in both genes, *csdE* and *csdF*, leads to a distinct decrease in total amount of Gal isolated from the LTA preparation (Figure 14). With this finding, a function for each of the identified GT gene pairs has been successfully assigned. Interestingly, each set of genes appears to glycosylate a distinct biochemical component of the bacterial cell envelope and/or cell wall. So far, similar three-component GT systems have been shown to be involved in the glycosylation of compounds such as the LPS-O antigens of Gram negative bacteria as well as WTA/LTA of Gram positive bacteria (39). Here, we present the first evidence for this system being employed by the *L. lactis* species as well as for its involvement in the glycosylation of lactococcal CWPS.

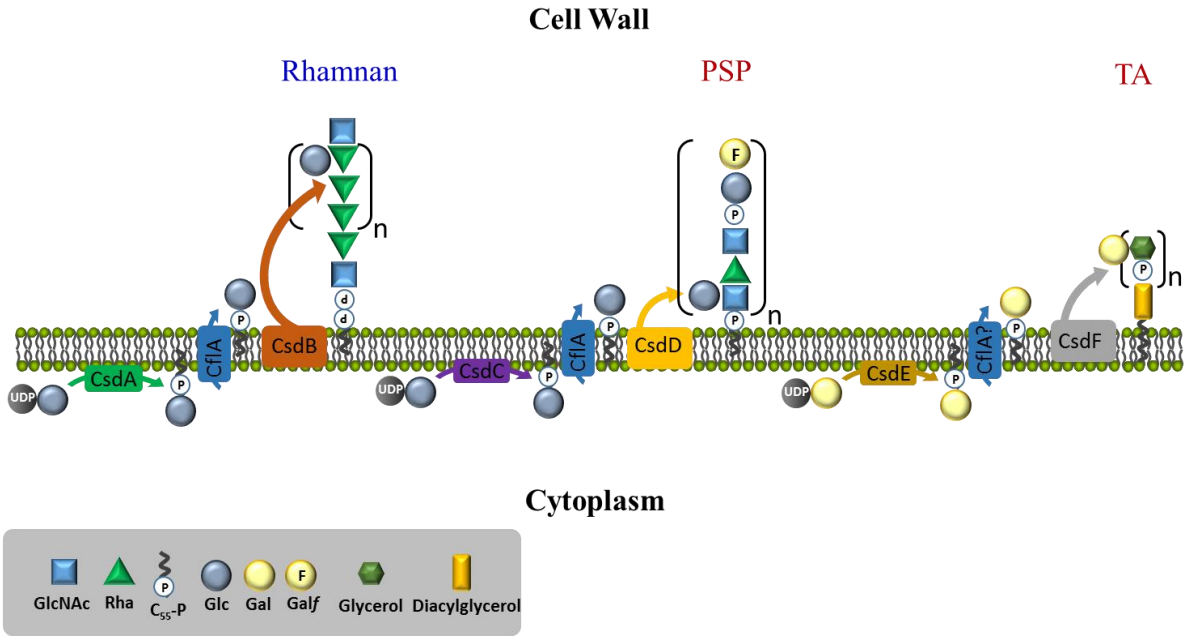


Figure 14. Schematic representation of the suggested functions of the proteins encode by each of the three identified gene pairs (*csdAB*, *csdCD*, *csdEF*) and the flippase-encoding gene *cflA*.

Furthermore, these data suggest that the PolM GT of each glucosylation system, CsdB or CsdD, is highly specific for a given cell wall glycopolymer (rhamnan or PSP, respectively) and that their optimal physiological functioning is likely in association with their cognate C_{55}-P Glc GT (Figure 14). In addition, in strains endowed with two glucosylation systems, such as the ones encoded by *csdAB* and *csdCD*, there appears to exist functional redundancy through a common pool of C_{55}-P-Glc intermediates. Of note, according to the model of the three-component mechanism involved in extracytoplasmic glycosylation of cell wall glycopolymers (39, 41), this lipid-sugar intermediate is synthesized in the cytoplasm and must be re-oriented to the other side of the membrane by a membrane flippase, such as CflA, in order to be used by the PolM GT. The observed interference between the two glucosylation systems suggests that the CflA may act as the third component of both glucosylation systems. This observation was further supported by the results presented here, whereby, the introduction of a non-sense mutation within *cflA* completely blocked the glucosylation of the rhamnan in *L. lactis* NZ9000. Together with the results presented in Chapter 2 of the thesis, it is clear that CflA forms part of both the CsdAB as well as the CsdCD glucosylation systems (Figure 14).

The *csdAB* and *csdCD* gene pairs present in *L. lactis* NZ9000 are both functional. However, their respective transcription levels (see also Chapter 2) and the activity of the encoded proteins differ quite dramatically. Indeed, as shown previously in Chapter 2, CsdCD glucosylates PSP at a high level, with (almost) all repeating units possessing a side chain Glc according to previous nuclear magnetic resonance (NMR) data (6). Thus *csdCD* is part of the biosynthesis pathway of PSP in *L. lactis* NZ9000 under standard (i.e. those employed in the laboratory) growth conditions. In contrast, *csdAB* exhibit low transcriptional activity in NZ9000 with a corresponding low level of Glc on rhamnan chains under such conditions. However, lactococcal strains that overproduce CsdAB can accommodate a high level of glucosylation on the rhamnan component of the CWPS. Whereas glucosylation of PSP, which

is exposed at the bacterial surface may primarily function in escaping phage predation, the role of glucosylation of rhamnan which is thought to be embedded into the cell wall peptidoglycan (9), remains unclear. However, preliminary data has implicated rhamnan glucosylation in phage-host interaction (covered in more detail below).

As mentioned above little is known about the importance of such glycosylation events in the CWPS structure of *L. lactis* and their apparent phenotypic effects on the species. However, it has previously been indicated that genes of similar amino acid homologies and functions in a number of species both Gram negative, such as *Shigella flexneri* (42), and Gram-positive, such as *Ls. monocytogenes* (38), play a particular role in the interaction of the bacterial strains and the phages that infect them. In more detail, a three-gene cluster present within bacteriophages commonly infecting *S. flexneri*, was shown to encode a glycosylation system that modulates the strain's serotype through the glucosylation of its O antigen, a structure that is required for bacteriophage adsorption. Such modification, limits further infection by closely related bacteriophages (42). Similarly, in the case of *Ls. monocytogenes*, the presence of a functionally active three-component glycosylation system provides the strain with phage resistance through the galactosylation of the phage receptor, which is known to be a WTA (38). Our results here corroborate these findings and extend the importance of such molecular mechanisms of bacteriophage defence to the lactococcal group. More specifically, overexpression of *csdAB* was shown to also act as a phage predation deterrent both in *L. lactis* 3107 (PSP(+) and rhamnan unexposed) and *L. lactis* VES5751 (PSP(-) and rhamnan exposed). Interestingly, the rhamnan glucose substitutions in *L. lactis* 3107 and VES751 appear to impact phage interaction in a divergent manner. In the case of *L. lactis* 3107, phage adsorption appears to be unaffected while in the case of the *L. lactis* VES5751 infection appears to be impaired at the level of phage adsorption onto the cell wall. This hints at the different modes of infection

employed by the two bacteriophages, in the case of the former a P335 lysogenic phage, LC3 (43), and in the case of the latter a 936 lytic phage, jj50* (44).

In addition to phage host interactions, cell wall glycopolymers and their substitutions have previously been shown to contribute to increased antibiotic resistance (45-47) and also contribute to biofilm formation (48-50). A recent study showed similarly that *Staphylococcus aureus* LTAs bear only a low level of glycosylation (GlcNAc substituents) under normal growth conditions, but this level is markedly raised when bacteria are subjected to salt-induced stress [35]. It is tempting to speculate that rhamnan glucosylation is also induced by environmental conditions or stresses, and could thus enable bacteria to adapt their cell wall physiology. Here, we investigated the effect of rhamnan as well as TA glycosylation on a number of phenotypic responses in *L. lactis*. Surprisingly, changes in the degree of glycosylation of either structure did not alter the strain's growth response in a number of stress inducing conditions, including increased salinity, decreased pH, added EDTA or ethanol. Interestingly, increased galactosylation of the LTA through the overexpression of *csdEF* led to the strain's greater biofilm forming ability, while increased glucosylation of the rhamnan by the overexpression of *csdAB* had the opposite effect. Finally, our results support previously found evidence for the mechanism of nisin resistance in *L. lactis* through increased teichoic-acid substitution and increased septum density (46). In detail, altering the levels of glycosylation of the teichoic acids in *L. lactis* NZ9000 led to equivalent changes in nisin resistance, i.e. increased glycosylation led to increased nisin resistance and *vice versa*.

Transcriptional analysis of the two gene pairs uncovered diametrically different levels and patterns of expression. The promoter of *csdAB* exhibited minimal levels of expression of the *lacZ* reporter gene and as a result could explain the already minimal levels of glycosylation seen on the wild-type rhamnan component of *L. lactis* NZ9000. However, the promoter of *csdEF* was subject to the highest levels of transcription of any of three gene pairs. Based on

our observation that *csdEF* is involved in LTA modification, we speculate that this abundance of *csdEF* throughout the cell's life cycle might be necessary due to the larger degree of sugar substituents found on LTAs in *L. lactis* relative to the sugar decoration of the CWPS and thus the greater need for glycosylation guided by *csdEF*.

These results, along with our findings presented in Chapter 2 of this thesis, shed further light on the mechanisms *L. lactis* employs for structurally modifying components of the bacterial cell wall. It was further shown that such alterations in structure could be linked to bacterial response to phage predation and perhaps bacteriocin resistance, however, further analysis and investigation should be performed to exactly pinpoint the phenotypic implications of such modifications of cell wall glycopolymers.

3.6 References

1. Song AAL, In LLA, Lim SHE, Rahim RA. 2017. A review on *Lactococcus lactis*: from food to factory. Microbial Cell Factories 16:15.
2. Cavanagh D, Fitzgerald GF, McAuliffe O. 2015. From field to fermentation: The origins of Lactococcus lactis and its domestication to the dairy environment. Food Microbiology 47:45-61.
3. Chapot-Chartier MP, Kulakauskas S. 2014. Cell wall structure and function in lactic acid bacteria. Microbial Cell Factories 13 Suppl 1:S9.
4. Delcour J, Ferain T, Deghorain M, Palumbo E, Hols P. 1999. The biosynthesis and functionality of the cell-wall of lactic acid bacteria. Antonie Van Leeuwenhoek 76:159-84.
5. Mistou MY, Sutcliffe IC, van Sorge NM. 2016. Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in gram-positive bacteria. FEMS Microbiology Reviews 40:464-79.
6. Chapot-Chartier MP, Vinogradov E, Sadovskaya I, Andre G, Mistou MY, Trieu-Cuot P, Furlan S, Bidnenko E, Courtin P, Pechoux C, Hols P, Dufrene YF, Kulakauskas S. 2010. Cell surface of *Lactococcus lactis* is covered by a protective polysaccharide pellicle. Journal of Biological Chemistry 285:10464-71.
7. Ainsworth S, Sadovskaya I, Vinogradov E, Courtin P, Guerardel Y, Mahony J, Grard T, Cambillau C, Chapot-Chartier M-P, van Sinderen D. 2014. Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. mBio 5:e00880-14.
8. Farenc C, Spinelli S, Vinogradov E, Tremblay D, Blangy S, Sadovskaya I, Moineau S, Cambillau C. 2014. Molecular insights on the recognition of a *Lactococcus lactis c*ell

wall pellicle by the phage 1358 receptor binding protein. Journal of Virology 88:7005-7015.

9. Sadovskaya I, Vinogradov E, Courtin P, Armalyte J, Meyrand M, Giaouris E, Palussière S, Furlan S, Péchoux C, Ainsworth S, Mahony J, van Sinderen D, Kulakauskas S, Guérardel Y, Chapot-Chartier M-P. 2017. Another brick in the wall: a rhamnan polysaccharide trapped inside peptidoglycan of *Lactococcus lactis*. mBio 8.

10. Mahony J, Cambillau C, van Sinderen D. 2017. Host recognition by lactic acid bacterial phages. FEMS Microbiology Reviews 41:S16-S26.

11. Dupont K, Janzen T, Vogensen FK, Josephsen J, Stuer-Lauridsen B. 2004. Identification of *Lactococcus lactis* genes required for bacteriophage adsorption. Applied and Environmental Microbiology 70:5825-32.

12. Mahony J, Kot W, Murphy J, Ainsworth S, Neve H, Hansen LH, Heller KJ, Sørensen SJ, Hammer K, Cambillau C, Vogensen FK, van Sinderen D. 2013. Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. Applied and Environmental Microbiology 79:4385-4392.

13. Vinogradov E, Sadovskaya I, Courtin P, Kulakauskas S, Grard T, Mahony J, van Sinderen D, Chapot-Chartier M-P. 2018. Determination of the cell wall polysaccharide and teichoic acid structures from *Lactococcus lactis* IL1403. Carbohydrate Research 462:39-44.

14. Vinogradov E, Sadovskaya I, Grard T, Murphy J, Mahony J, Chapot-Chartier M-P, van Sinderen D. 2018. Structural studies of the cell wall polysaccharide from *Lactococcus lactis* UC509.9. Carbohydrate Research 461:25-31.

15. Kuipers OP, de Ruyter PGGA, Kleerebezem M, de Vos WM. 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. Journal of Biotechnology 64:15-21.

16. Braun V, Hertwig S, Neve H, Geis A, Teuber M. 1989. Taxonomic differentiation of bacteriophages of *Lactococcus lactis* by electron microscopy, DNA-DNA hybridization, and protein profiles. Microbiology 135:2551-2560.

17. van Pijkeren JP, Britton RA. 2012. High efficiency recombineering in lactic acid bacteria. Nucleic Acids Research 40:e76.

18. McGrath S, Fitzgerald GF, van Sinderen D. 2001. Improvement and optimization of two engineered phage resistance mechanisms in *Lactococcus lactis*. Applied and Environmental Microbiology 67:608-16.

19. O'Driscoll J, Glynn F, Cahalane O, O'Connell-Motherway M, Fitzgerald GF, van Sinderen D. 2004. Lactococcal plasmid pNP40 encodes a novel, temperature-sensitive restriction-modification system. Applied and Environmental Microbiology 70:5546-5556.

20. Higgins DL, Sanozky-Dawes RB, Klaenhammer TR. 1988. Restriction and modification activities from *Streptococcus lactis* ME2 are encoded by a self-transmissible plasmid, pTN20, that forms cointegrates during mobilization of lactose-fermenting ability. Journal of Bacteriology 170:3435-3442.

21. Chandry PS, Moore SC, Boyce JD, Davidson BE, Hillier AJ. 1997. Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. Molecular Microbiology 26:49-64.

22. Lillehaug D, Lindqvist B, Birkeland NK. 1991. Characterization of phiLC3, a *Lactococcus lactis* subsp. *cremoris* temperature bacteriophage with cohesive single-stranded DNA ends. Applied and Environmental Microbiology 57:3206-3211.

23. Oliveira J, Mahony J, Hanemaaijer L, Kouwen TRHM, van Sinderen D. 2018. Biodiversity of bacteriophages infecting *Lactococcus lactis* starter cultures. Journal of Dairy Science 101:96-105.

24. Oliveira J, Mahony J, Lugli GA, Hanemaaijer L, Kouwen T, Ventura M, van Sinderen D. 2016. Genome Sequences of Eight Prophages Isolated from Lactococcus lactis Dairy Strains. Genome Announcements 4.

25. Oliveira J, Mahony J, Hanemaaijer L, Kouwen TRHM, Neve H, MacSharry J, van Sinderen D. 2017. Detecting *Lactococcus lactis* prophages by mitomycin C-mediated induction coupled to flow cytometry analysis. Frontiers in Microbiology 8:1-11.

26. Sijtsma L, Wouters JT, Hellingwerf KJ. 1990. Isolation and characterization of lipoteichoic acid, a cell envelope component involved in preventing phage adsorption, from *Lactococcus lactis* subsp. *cremoris* SK110. Journal of bacteriology 172:7126-7130.

27. Ciucanu I, Kerek F. 1984. A simple and rapid method for the permethylation of carbohydrates. Carbohydrate Research 131:209-217.

28. Read SM, Currie G, Bacic A. 1996. Analysis of the structural heterogeneity of laminarin by electrospray-ionisation-mass spectrometry. Carbohydrate Research 281:187-201.

29. Lindberg B, Lonngren J. 1978. Methylation analysis of complex carbohydrates: general procedure and application for sequence analysis. Methods in Enzymology 50:3-33.

30. Mahony J, McGrath S, Fitzgerald GF, van Sinderen D. 2008. Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. Applied and Environmental Microbiology 74:6206-6215.

31. Lillehaug D. 1997. An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. Journal of Applied Microbiology 83:85-90.

32. Østergaard Breum S, Neve H, Heller KJ, Vogensen FK. 2007. Temperate phages TP901-1 and φLC3, belonging to the P335 species, apparently use different pathways

for DNA injection in *Lactococcus lactis* subsp. *cremoris* 3107. FEMS Microbiology Letters 276:156-164.

33. Stockdale SR, Collins B, Spinelli S, Douillard FP, Mahony J, Cambillau C, van Sinderen D. 2015. Structure and assembly of TP901-1 virion unveiled by mutagenesis. PLOS ONE 10:e0131676.

34. Van Pijkeren JP, Neoh KM, Sirias D, Findley AS, Britton RA. 2012. Exploring optimization parameters to increase ssDNA recombineering in *Lactococcus lactis* and *Lactobacillus reuteri*. Bioengineered 3:209-17.

35. Field D, Gaudin N, Lyons F, O'Connor PM, Cotter PD, Hill C, Ross RP. 2015. A bioengineered nisin derivative to control biofilms of *Staphylococcus pseudintermedius*. PLOS ONE 10:e0119684.

36. Vázquez-Sánchez D, Cabo ML, Ibusquiza PS, Rodríguez-Herrera JJ. 2014. Biofilm-forming ability and resistance to industrial disinfectants of *Staphylococcus aureus* isolated from fishery products. Food Control 39:8-16.

37. Rush JS, Edgar RJ, Deng P, Chen J, Zhu H, van Sorge NM, Morris AJ, Korotkov KV, Korotkova N. 2017. The molecular mechanism of N-acetylglucosamine side-chain attachment to the Lancefield group A Carbohydrate in *Streptococcus pyogenes*. The Journal of Biological Chemistry 292:19441-19457.

38. Spears PA, Havell EA, Hamrick TS, Goforth JB, Levine AL, Abraham ST, Heiss C, Azadi P, Orndorff PE. 2016. *Listeria monocytogenes* wall teichoic acid decoration in virulence and cell-to-cell spread. Molecular Microbiology 101:714-30.

39. Mann E, Whitfield C. 2016. A widespread three-component mechanism for the periplasmic modification of bacterial glycoconjugates. Canadian Journal of Chemistry 94:883-893.

40. Kramer NE, Hasper HE, van den Bogaard PTC, Morath S, de Kruijff B, Hartung T, Smid EJ, Breukink E, Kok J, Kuipers OP. 2008. Increased d-alanylation of lipoteichoic acid and a thickened septum are main determinants in the nisin resistance mechanism of Lactococcus lactis. Microbiology 154:1755-1762.

41. Kho K, Meredith TC. 2018. Salt-induced stress stimulates a lipoteichoic acid-specific three-component glycosylation system in *Stsaphylococcus aureus*. Journal of Bacteriology 200.

42. Guan S, Bastin DA, Verma NK. 1999. Functional analysis of the O antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. Microbiology 145:1263-1273.

43. Mahony J, Oliveira J, Collins B, Hanemaaijer L, Lugli GA, Neve H, Ventura M, Kouwen TR, Cambillau C, van Sinderen D. 2017. Genetic and functional characterisation of the lactococcal P335 phage-host interactions. BMC Genomics 18:146.

44. Dupont K, Vogensen FK, Neve H, Bresciani J, Josephsen J. 2004. Identification of the Receptor-Binding Protein in 936-Species Lactococcal Bacteriophages. Applied and Environmental Microbiology 70:5818-5824.

45. Saar-Dover R, Bitler A, Nezer R, Shmuel-Galia L, Firon A, Shimoni E, Trieu-Cuot P, Shai Y. 2012. D-Alanylation of lipoteichoic acids confers resistance to cationic peptides in Group B *Streptococcus* by increasing the cell wall density. PLOS Pathogens 8:e1002891.

46. Kramer NE, Hasper HE, van den Bogaard PT, Morath S, de Kruijff B, Hartung T, Smid EJ, Breukink E, Kok J, Kuipers OP. 2008. Increased D-alanylation of lipoteichoic acid and a thickened septum are main determinants in the nisin resistance mechanism of *Lactococcus lactis*. Microbiology 154:1755-62.

47. Giaouris E, Briandet R, Meyrand M, Courtin P, Chapot-Chartier M-P. 2008. Variations in the degree of D-alanylation of teichoic acids in *Lactococcus lactis* alter resistance to cationic antimicrobials but have no effect on bacterial surface hydrophobicity and charge. Applied and Environmental Microbiology 74:4764-4767.

48. Vu B, Chen M, Crawford R, Ivanova E. 2009. Bacterial extracellular polysaccharides involved in biofilm formation. Molecules 14:2535.

49. Vergara-Irigaray M, Maira-Litrán T, Merino N, Pier GB, Penadés JR, Lasa I. 2008. Wall teichoic acids are dispensable for anchoring the PNAG exopolysaccharide to the *Staphylococcus aureus* cell surface. Microbiology 154:865-877.

50. Vinogradov E, Sadovskaya I, Li J, Jabbouri S. 2006. Structural elucidation of the extracellular and cell-wall teichoic acids of *Staphylococcus aureus* MN8m, a biofilm forming strain. Carbohydrate Research 341:738-43.

Chapter 4

A dual chain assembly mechanism to synthesize a complex cell wall rhamnopolysaccharide with high structural diversity in *Lactococcus lactis*

NB. CWPS Structural Characterization and Analysis performed by the group of Dr. Marie-Pierre Chapot-Chartier, French National Institute of Agricultural Research (INRA), Jouy-en-Josas, France, while primer extension analysis was performed by Dr. Christophe Penno, University College Cork, Ireland

Manuscript in Preparation

Table of Contents

4.1 Abstract 157

4.2 Introduction 158

4.3 Methods 161

4.3.1 Strains phages and growth conditions Conditions 161

4.3.2 Promoter mapping and reverse transcription PCR 163

4.3.3 Primer extension analysis and Reverse Transcription PCR 171

4.3.4 CRISPR-Recombineering 172

4.3.5 CWPS structural determination 175

4.3.5 Determination of fitness of *cwps* mutants 175

4.3.6 Bacteriophage isolation and assays 176

4.4 Results 177

4.4.1 Promoter mapping of the *cwps* gene cluster 177

4.4.2 Transcription start site identification and RT-PCR 179

4.4.3 *L. lactis* NZ9000 *cwps* cluster non-sense mutations and CWPS structural analysis 183

4.4.4 Design of a tentative model for PSP biosynthesis and attachment to rhamnan 187

4.4.5 Assessment of the model by phenotypic analysis of the *cwps* mutants. 191

4.5 Discussion 195

4.6 References 203

4.1 Abstract

The genetic components responsible for the assembly of the cell wall-associated polysaccharide (CWPS) in *Lactococcus lactis* have been mapped to a large gene cluster, being 25-30 kb in length, and containing over 20 open reading frames (ORFs). Despite conclusive evidence that gene content and genetic diversity within this cluster is directly linked to the final CWPS structure produced by each *L. lactis* strain, little knowledge exist on the functions of the individual proteins encoded by this cluster or how this cluster is transcribed. A model has previously been put forward that outlines a molecular assembly pathway for a subcomponent of the *L. lactis* CWPS, known as the rhamnan. Through directed non-sense mutations targeted at individual *cwps* genes (*wpsABCDEFHIJ*) and structural analysis of the CWPS produced by such mutant strains of *L. lactis* NZ9000, we propose a full biochemical assembly pathway for the complete CWPS, incorporating its two constituents PSP and rhamnan. We also provide transcriptional analysis and phage/host interaction data that corroborate our proposed CWPS biosynthesis model.

4.2 Introduction

Lactococcus lactis is a bacterial species of tremendous economic importance due to its extensive application in the context of dairy fermentations (1). Similarly, *L. lactis* has increasingly been promoted as an alternative production host to yeast and *Escherichia coli* for large-scale production and/or *in situ* delivery of metabolites and bioactive compounds (2, 3). Consequently, extensive research has focused on extending knowledge of the physiology and functionality of this species (4-7). A unique aspect of lactococcal physiology that has received continued interest in recent years represents the molecular components of the *L. lactis* cell envelope. Seminal work carried out by Chapot-Chartier *et al.* (8) uncovered a unique component of the cell wall polysaccharide (CWPS) structure in *L. lactis* MG1363, known as PSP. This structure represents a thin, yet compact outer layer of the bacterial cell envelope and is composed of phosphate-containing hexasaccharide repeating units that consist of the saccharidic residues rhamnose (Rha), glucose (Glc), N-acetylglucosamine (GlcNAc), and galactofuranose (Gal*f*) (8). Similar, yet distinct in their composition, PSP structures have been identified in other *L. lactis* strains, including *L. lactis* 3107 (9) and SMQ-388 (10).

More recently, it was shown that both *L. lactis* MG1363 and 3107 contain an additional neutral, rhamnose-containing polysaccharide, or rhamnan, which together with the PSP forms the CWPS (11). Following genomic analysis of PSP-negative variants of *L. lactis* MG1363 (8), the molecular machinery responsible for assembling the PSP was mapped to a gene cluster (called the *cwps* gene cluster), approximately 25 kb in length and containing 22 open reading frames (ORFs). Such a *cwps* gene cluster is present in the genomes of all *L. lactis* strains that have been sequenced to date, albeit with a large degree of sequence and gene content diversity, particularly at the 3' end of *cwps* (9, 12). The more conserved 5' portion of the gene cluster is proposed to encode the biosynthetic abilities for the production of the rhamnan component of

the CWPS, although the two structures are believed to be produced independently of each other since PSP-negative variants still retain the rhamnan layer in their cell envelope (11).

Based on sequence identity, *L. lactis* strains have been classified into three distinct *cwps* genotypes, i.e. types A, B, and C (12). Furthermore, five distinct subtypes of the C-type strains were also identified, designated subtype C_1 through to C_5, based on the diversity of glycosyltransferase (GT)-encoding genes (9). It has been verified that the observed *cwps* sequence diversity corresponds to structural differences of the associated CWPS. Mutating the variable GT-encoding genes of the C_1 subtype strain *L. lactis* NZ9000, and introducing GT-encoding genes of a C_2 subtype strain (3107) in the resulting mutant, caused conversion of the CWPS composition from C_1 to the C_2 subtype (9). More recently, the CWPS structure of both an A-type strain, *L. lactis* UC509.9, and a B-type strain, *L. lactis* IL1403, were biochemically characterized (13, 14). The CWPS of *L. lactis* UC509.9 is composed of a linear backbone of a repeating tetrasaccharide subunit, which contains Rha and Glc, with regular substitutions of a phosphorylated oligosaccharide made up of a common trisaccharide along with three non-stoichiometric substitutions (13). The CWPS of *L. lactis* IL1403 is composed of a glycan polymer consisting of a linear Rha disaccharide subunit, being regularly decorated with a trisaccharide branch, which in turn may contain a glycerophosphate group substitution (14). Since the CWPS, and in particular its decorative branch (such as the PSP), has been shown to act as the receptor for many phages, it follows that the observed differences in CWPS structure are responsible, at least partially, for the high degree of specificity of lactococcal phages and their observed narrow host range (9, 10, 12).

While it is established that the *cwps* gene cluster is responsible for CWPS biosynthesis including both the rhamnan and PSP constituents, functional assignment is not available at present for every protein encoded by the gene cluster. Previously, a model for rhamnan biosynthesis in *L. lactis* NZ9000, involving an ABC transporter-dependent pathway was

proposed (11) assigning a function to nine proteins encoded in the 5’ part of the *cwps* cluster. In contrast, little is known about the biosynthesis of the PSP repeat units, their export and polymerization.Based on transcriptional and mutational analyses of the *cwps* gene cluster, combined with growth assessment of *cwps* mutants and phage escape mutant isolation and characterisation, we propose a comprehensive model for CWPS biosynthesis, based on a dual assembly mechanism for its constituents PSP and rhamnan.

4.3 Methods

4.3.1 Strains phages and growth conditions Conditions

Bacterial strains used in this study are listed in Table 1. Strains were grown at 30 ^{o}C overnight in M17 broth and/or M17 agar (Oxoid Ltd, Hampshire, United Kingdom) supplemented with 5 g glucose per liter or kg of M17 medium. Chloramphenicol, tetracycline, or erythromycin (Sigma-Aldrich, Missouri, USA) was added to media, where appropriate, to a final concentration of 5 μg/ml, 10 μg/ml, or 5 μg/ml, respectively. For controlled transcription of genes placed under the nisin-inducible promoter, P^{nisA}, nisaplin (DuPont, Copenhagen, Denmark) was supplemented to the growth medium at a final concentration of 40 ng/ml.

Table 1. Strains, plasmids, bacteriophages used, constructed and/or isolated in this study

Strain, plasmid, or phage	Feature(s)	Source
Bacterial Strains		
L. lactis subsp. *cremoris* NZ9000	*L. lactis* MG1363 derivative containing *nisRK*, host to phages jj50, p2, and sk1	(15)
L. lactis subsp. *cremoris* VES5751	*L. lactis* MG1363 derivative exhibiting a deficient PSP phenotype due to a mutation in *llmg_0226*	(8)
CWPS Mutant Strain		
L. lactis subsp. *cremoris* NZ9000-*wpsJ*	NZ9000 with a TGATAACCC inserted in *llnz_01120* resulting in the introduction of a TGA and a TAA stop codon in this gene	This work
L. lactis subsp. *cremoris* NZ9000-*wpsA*	NZ9000 with a TGATAACCC inserted in *llnz_01135* resulting in the introduction of a TGA and a TAA stop codon in this gene	This work
L. lactis subsp. *cremoris* NZ9000-*wpsB*	NZ9000 with a TGATAACCC inserted in *llnz_01140* resulting in the introduction of a TGA and a TAA stop codon in this gene	This work
L. lactis subsp. *cremoris* NZ9000-*wpsC*	NZ9000 with a GAATTC inserted in *llnz_01145* resulting in the introduction of a TGA stop codon in this gene	(9)
L. lactis subsp. *cremoris* NZ9000-*wpsD*	NZ9000 with a TAATGACCC inserted in *llnz_01150* resulting in the introduction of a TAA and a TGA stop codon in this gene	This work
L. lactis subsp. *cremoris* NZ9000-*wpsE*	NZ9000 with a TGATAACCC inserted in *llnz_01155* resulting in the introduction of a TGA and a TAA stop codon in this gene	This work
L. lactis subsp. *cremoris* NZ9000-*wpsF*	NZ9000 with a TGATAACCC inserted in *llnz_01160* resulting in the introduction of a TGA and a TAA stop codon in this gene	This work
L. lactis subsp. *cremoris* NZ9000-*wpsH*	NZ9000 with a GATATCG inserted in *llnz_01175* resulting in the intriduction of a TGA stop codon insertion in this gene	This work
L. lactis subsp. *cremoris* NZ9000-*wpsI*	NZ9000 with a TGATAACCC inserted in *llnz_01160* resulting in the introduction of a TGA and a TAA stop codon in this gene	This work
Plasmids		
pCNR	Recombineering-facilitating vector containing *recT*, P*nisA*, Cm^r derived from the low-copy vector pPTPi	This work

pVPL3004	Low-copy vector expressing *cas9* along with tracRNA, Ery^r	(16)
pCRISPR	High-copy vector carrying CRISPR repeats and used for integrating targeting spacer sequences, Tet^r	(16)
pPTPL	*E. coli-L. lactis* promoter-probe vector, Tet^r	(17)
pPTPL::NZProm1-8	pPTPL containing one of the 9 intergenic regions found within the NZ9000 CWPS gene cluster	This work
Bacteriophage		
jj50	936 species, propagated on NZ9000	(18)
sk1	936 species, propagated on NZ9000	(19)
p2	936 species, propagated on NZ9000	(20)
MCC1	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	Isolated by S. Kulakauskas, INRA Collection
MCC5	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	Isolated by S. Kulakauskas, INRA Collection
MCC17	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	Isolated by S. Kulakauskas, INRA Collection
IT1	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT2	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT3	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT4	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT5	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work

4.3.2 Promoter mapping and reverse transcription PCR

All recombinant plasmids (Table 1) were generated in *L. lactis* NZ9000 and the primers (Table 2), unless otherwise indicated, were ordered from Eurofins MWG (Ebersberg, Germany). All intergenic regions present within the *cwps* gene cluster of strains *L. lactis* NZ9000, defined as the nucleotide sequence directly upstream of any annotated gene start codon and directly

downstream of the stop codon of the corresponding upstream gene (> 20 bp), were amplified using the oligonucleotides listed in Table 2 with Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Massachusetts, USA) and cloned into the low copy, promoter-probe vector pPTPL (17). The resulting constructs are listed in Table 1. Promoter activity was assessed visually by plating the *L. lactis* NZ9000 harbouring the above-mentioned constructs onto GM17 agar plates containing 10 μg/ml X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, Sigma-Aldrich, USA), as well as quantitatively by measuring their specific β-galactosidase activity during early lag, mid-exponential, and late-exponential growth, as described previously using the formula $1000 \times [(OD_{420nm} - 1.75 \times OD_{550nm})] / (T \times V \times OD_{600nm})$, where T represents the total time of the reaction in minutes and V represents the total volume of the culture, in ml, used in the assay (21).

Table 2. Primers used in the development of the promoter-probe pPTPL and pCRISPR constructs, for screening of *cwps* mutants, and primer extension analysis of identified *cwps* promoter regions.

Purpose	Sequence 5'-3'	Oligo Name	pPTPL Construct
Fwd primer for cloning intergenic region upstream of *llnz_01075* into pPTPL	gcgcgcggatcccagtaacgattttctctgg	oIT95	*NZProm1*
Rev primer for cloning intergenic region upstream of *llnz_01075* into pPTPL	gcgcgctctagacagttgttttgacgcagc	oIT96	*NZProm1*
Fwd primer for cloning intergenic region upstream of *llnz_01085* into pPTPL	gcgcgcggatcctggtcttcaatgggaagt	oIT97	*NZProm2*
Rev primer for cloning intergenic region upstream of *llnz_01085* into pPTPL	gcgcgctctagattcgacacgtccgctgtca	oIT98	*NZProm2*
Fwd primer for cloning intergenic region upstream of *llnz_01090* into pPTPL	gcgcgcggatcccagaacttggttggactc	oIT99	*NZProm3*
Rev primer for cloning intergenic region upstream of *llnz_01090* into pPTPL	gcgcgctctagagacgcaactctgttccaa	oIT100	*NZProm3*
Fwd primer for cloning intergenic region upstream of *llnz_01095* into pPTPL	gcgcgcggatccaagcgacaggttttgtca	oIT101	*NZProm4*
Rev primer for cloning intergenic region upstream of *llnz_01095* into pPTPL	gcgcgctctagacaaaccctccatatttcgc	oIT102	*NZProm4*
Fwd primer for cloning intergenic region upstream of *llnz_01135* into pPTPL	gcgcgcggatccaacggacctaacacaaa cg	oIT103	*NZProm5*
Rev primer for cloning intergenic region upstream of *llnz_01135* into pPTPL	gcgcgctctagattgataaggacaactggc	oIT104	*NZProm5*
Fwd primer for cloning intergenic region upstream of *llnz_01145* into pPTPL	gcgcgcggatccatttgggtacaagtattgc	oIT105	*NZProm6*
Rev primer for cloning intergenic region upstream of *llnz_01145* into pPTPL	gcgcgctctagaccacttctgatattccttc	oIT106	*NZProm6*
Fwd primer for cloning intergenic region upstream of *llnz_01160* into pPTPL	gcgcgcggatccctaagagctactggtaag t	oIT107	*NZProm7*
Rev primer for cloning intergenic region upstream of *llnz_01160* into pPTPL	gcgcgctctagatgtctgaatatggtgtgag	oIT108	*NZProm7*
Fwd primer for cloning intergenic region upstream of *llnz_01175* into pPTPL	gcgcgcggatccatctggtgcagtgatgtt	oIT109	*NZProm8*
Rev primer for cloning intergenic region upstream of *llnz_01175* into pPTPL	gcgcgctctagagccccatcactccaataa	oIT110	*NZProm8*

Fwd primer for cloning intergenic region upstream of *rmlA* into pPTPL	aaaaaagaattccttggttgctcaaccaagaag	oIT111	*ILProm1*
Rev primer for cloning intergenic region upstream of *rmlA* into pPTPL	aaaaaatctagatgattgtccttcttgtactttg	oIT112	*ILProm1*
Fwd primer for cloning intergenic region upstream of *ybjD* into pPTPL	aaaaaagaattcgaaattgcttatcgtatggg	oIT113	*ILProm2*
Rev primer for cloning intergenic region upstream of *ybjD* into pPTPL	aaaaaatctagatgattgtccttcttgtactttg	oIT114	*ILProm2*
Fwd primer for cloning intergenic region upstream of *cpsM* into pPTPL	aaaaaagaattcgatgtacctaataacctaacg	oIT115	*ILProm3*
Rev primer for cloning intergenic region upstream of *cpsM* into pPTPL	aaaaaatctagaatttttctttctaatcactatttc	oIT116	*ILProm3*
Fwd primer for cloning intergenic region upstream of *rmlB* into pPTPL	aaaaaagaattcaaatgtcggtggtttactc	oIT117	*ILProm4*
Rev primer for cloning intergenic region upstream of *rmlB* into pPTPL	aaaaaatctagagtttttttccttttttgattatt	oIT118	*ILProm4*
Fwd primer for cloning intergenic region upstream of *rmlC* into pPTPL	aaaaaagaattcgtgaaaacgaaaactggtg	oIT119	*ILProm5*
Rev primer for cloning intergenic region upstream of *rmlC* into pPTPL	aaaaaatctagattgttctccaatctttaatttttt	oIT120	*ILProm5*
Fwd primer for cloning intergenic region upstream of *rgpA* into pPTPL	aaaaaaggattcgactctgtcactaaaataaacta	oIT121	*ILProm6*
Rev primer for cloning intergenic region upstream of *rgpA* into pPTPL	aaaaaatctagagttgttccttttacttactaaa	oIT122	*ILProm6*
Fwd primer for cloning intergenic region upstream of *ycaG* into pPTPL	aaaaaagaattcaaagctattggaattg	oIT123	*ILProm7*
Rev primer for cloning intergenic region upstream of *ycaG* into pPTPL	aaaaaatctagagaagtccaaacttccgaaa	oIT124	*ILProm7*
Fwd primer for cloning intergenic region upstream of *ycbB* into pPTPL	aaaaaagaattcgacgaaagcagaattactg	oIT125	*ILProm8*
Rev primer for cloning intergenic region upstream of *ycbB* into pPTPL	aaaaaatctagaaaatctcttttccctaattaaac	oIT126	*ILProm8*
Fwd primer for cloning intergenic region upstream of *ycbJ* into pPTPL	aaaaaagaattccttctttacatagcttatgg	oIT127	*ILProm9*
Rev primer for cloning intergenic region upstream of *ycbJ* into pPTPL	aaaaaatctagaaatttatattttattttccatatta	oIT128	*ILProm9*

Fwd oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01120*	aaacacgtctttttttaatgattgctcttgttgcg	oIT45	N/A
Rev oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01120*	aaaacgcaacaagagcaatcattaaaaaaagacgt	oIT46	N/A
Fwd primer for sequencing and screening for mutated variants of *llnz_01120*	gcagcacaagaagatagcag	oIT47	N/A
Rev primer for sequencing and screening for mutated variants of *llnz_01120*	gcagtcatagcttaaacttgtagc	oIT48	N/A
Fwd primer annealing to the inserted sequence in *llnz_01120* used for screening mutant variants	gctcttgtttgataaccc	oIT49	N/A
Fwd oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_1135*	aaacagggaatattagaaaatctaaaaattgtagg	oIT50	N/A
Rev oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01135*	aaaacctacaatttttagattttctaatattccct	oIT51	N/A
Fwd primer for sequencing and screening for mutated variants of *llnz_01135*	acaatgatttggtatcgtccac	oIT52	N/A
Rev primer for sequencing and screening for mutated variants of *llnz_01135*	cacccacctctttgacacg	oIT53	N/A
Rev primer annealing to the inserted sequence in *llnz_01135* used for screening mutant variants	ggcttgatagggttatca	oIT54	N/A
Fwd oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01140*	aaaccagctttcttctttatctatacgattcgttg	oIT55	N/A
Rev oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01140*	aaaacaacgaatcgtatagataaagaagaaagctg	oIT56	N/A
Fwd primer for sequencing and screening for mutated variants of *llnz_01140*	ccttaccaaacttgattgaacc	oIT57	N/A
Rev primer for sequencing and screening for mutated variants of *llnz_01140*	caatgcctccaccttcac	oIT58	N/A
Fwd primer annealing to the inserted sequence in *llnz_01140* used for screening mutant variants	cgattcgttgataaccc	oIT59	N/A
Fwd oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01150*	aaacattcaattttagaacaaacatacaaaaactg	oIT60	N/A

Rev oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01150*	aaaacagtttttgtatgtttgttctaaaattgaat	oIT61	N/A
Fwd primer for sequencing and screening for mutated variants of *llnz_01150*	ttatgaaattgatgaatactctaaacctag	oIT62	N/A
Rev primer for sequencing and screening for mutated variants of *llnz_01150*	cagcaatagataggacttgaag	oIT63	N/A
Fwd primer annealing to the inserted sequence in *llnz_01150* used for screening mutant variants	catacaaaaactaatgaccc	oIT64	N/A
Fwd oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01155*	aaacgtaacatacttggatagcgaagatacagatg	oIT65	N/A
Rev oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01155*	aaaacatctgtatcttcgctatccaagtatgttac	oIT66	N/A
Fwd primer for sequencing and screening for mutated variants of *llnz_01155*	gatatatcaatgtatgaagatgc	oIT67	N/A
Rev primer for sequencing and screening for mutated variants of *llnz_01155*	cttcaagactatttagaacc	oIT68	N/A
Fwd primer annealing to the inserted sequence in *llnz_01155* used for screening mutant variants	gaagatacatgataaccc	oIT69	N/A
Fwd oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01160*	aaacaatagactcagacactatatttaaggatagg	oIT70	N/A
Rev oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01160*	aaaacctatccttaaatatagtgtctgagtctatt	oIT71	N/A
Fwd primer for sequencing and screening for mutated variants of *llnz_01160*	acaaatttatgggcttttacc	oIT72	N/A
Rev primer for sequencing and screening for mutated variants of *llnz_01160*	cctggtccataaagattg	oIT73	N/A
Fwd primer annealing to the inserted sequence in *llnz_01160* used for screening mutant variants	tatttaaggattgataaccc	oIT74	N/A
Fwd primer for sequencing and screening for mutated variants of *llnz_01175*	cagtaccttcttccattaaatcgg	oIT75	N/A
Rev primer for sequencing and screening for mutated variants of *llnz_01175*	actttaacttcacccatatctgg	oIT76	N/A

Fwd primer annealing to the inserted sequence in *llnz_01175* used for screening mutant variants	tgtaagtgcttgatatcg	oIT77	N/A
Fwd oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01180*	aaactgctcttttctttggaattatcctatccatg	oIT78	N/A
Rev oligonucleotide for cloning CRISPR spacer into pCRISPR targeting the mutated region of *llnz_01180*	aaaacatggataggataattccaaagaaaagagca	oIT79	N/A
Fwd primer for sequencing and screening for mutated variants of *llnz_01180*	ggtttgggagaaggaaaag	oIT80	N/A
Rev primer for sequencing and screening for mutated variants of *llnz_01180*	cacagctacaagaaaagc	oIT81	N/A
Fwd primer annealing to the inserted sequence in *llnz_01160* used for screening mutant variants	atcctatcctgataaccc	oIT82	N/A
Amplification of *NZProm1* promoter fragments with IRD700-labelled oligonucleotides	catgaaagggattattttagcg	oIT83	N/A
Amplification of *NZProm1* promoter fragments with IRD700-labelled oligonucleotides	ctttacccactgacacgtgc	oIT84	N/A
Fwd primer, amplification of region containing *NZProm1* promoter region for sequencing ladders	aaacgtcaatttaacatccaatc	oIT85	N/A
Rev primer, amplification of region containing *NZProm1* promoter region for sequencing ladders	gcacgtgtcagtgggtaaag	oIT86	N/A
Amplification of *NZProm5* promoter fragments with IRD700-labelled oligonucleotides	cgatgacaaaaaaatgaaaattttac	oIT87	N/A
Amplification of *NZProm5* promoter fragments with IRD700-labelled oligonucleotides	gaaaatctaaaaattgtagagga	oIT88	N/A
Fwd primer, amplification of region containing *NZProm5* promoter region for sequencing ladders	gtaattatcgcctaccgtttg	oIT89	N/A
Rev primer, amplification of region containing *NZProm5* promoter region for sequencing ladders	tcctctacaatttttagattttc	oIT90	N/A
Amplification of *NZProm8* promoter fragments with IRD700-labelled oligonucleotides	gtataaaaaggtaaccttcttag	oIT91	N/A
Amplification of *NZProm8* promoter fragments with IRD700-labelled oligonucleotides	cttagttacattggagtgatgg	oIT92	N/A

Fwd primer, amplification of region containing *NZProm8* promoter region for sequencing ladders	caattctttgaagatgaatatgc	oIT93	N/A
Rev primer, amplification of region containing *NZProm8* promoter region for sequencing ladders	ccatcactccaatgtaactaag	oIT94	N/A
Fwd primer, amplification of cDNA targeting the presumed rhamnan transcriptional sub-unit	caacctacaacggcgaaaagt	oIT129	N/A
Rev primer, amplification of cDNA targeting the presumed rhamnan transcriptional sub-unit	gataccaatcatgcatgataatg	oIT130	N/A
Fwd primer, amplification of cDNA targeting the presumed PSP transcriptional sub-unit	gaagagatgaaaatggcggta	oIT131	N/A
Rev primer, amplification of cDNA targeting the presumed PSP transcriptional sub-unit	ccatataatttagcaactggtccag	oIT132	N/A
Fwd primer, amplification of cDNA targeting the presumed PSP-polymerization transcriptional sub-unit	gaggttcattttagtgctctg	oIT133	N/A
Rev primer, amplification of cDNA targeting the presumed PSP-polymerization transcriptional sub-unit	ttaaagaacggcattccaatgg	oIT134	N/A
Fwd primer, amplication of cDNA targeting the region surrounding the *NZProm5* region	ggcttgatgattgcggacgtt	oIT135	N/A
Rev primer, amplication of cDNA targeting the region surrounding the *NZProm5* region	ctgcgcgataaccactagtaac	oIT136	N/A
Fwd primer, amplication of cDNA targeting the region surrounding the *NZProm8* region	cactttggccatatttgcgca	oIT137	N/A
Rev primer, amplication of cDNA targeting the region surrounding the *NZProm8* region	ctgagaaatcacattcgccgt	oIT138	N/A
Fwd primer, amplication of cDNA targeting the region directly upstream of *cwps* gene cluster	ctttagacattatatttacctccc	oIT139	N/A
Rev primer, amplication of cDNA targeting the region directly upstream of *cwps* gene cluster	ccaacatcaaagttgacaatgg	oIT140	N/A

Table 3. Mutational oligonucleotides used in the development of the cwps gene mutant derivatives using CRISPR-Recombineering.

Purpose	Sequence 5'-3'	Oligo Name
Recombineering Oligo used in the stop codon insertion into *llnz_01120*	t*c*t*g*g*ttcaataaagagtggttgagcaactgtaaaaattaaaccgggttatcaaacaagagcaatcattaaaaaaagacgttctggtttg	oIT143
Recombineering Oligo used in the stop codon insertion into *llnz_01135*	t*t*t*a*t*cactacatagtctaattgataaggacaactggcttgatagggttatcaaatttttagattttctaatattccctcaccttcattg	oIT144
Recombineering Oligo used in the stop codon insertion into *llnz_01140*	a*a*c*c*a*ttttaacatatgtcttatttctgctctatctttacggggttatcaacgaatcgtatagataaagaagaaagctgccaaaataatt	oIT145
Recombineering Oligo used in the stop codon insertion into *llnz_01150*	g*g*t*a*g*aaccatcatcaataagtataatttccaagtttttgtcgaattcatgtttgttctaaaattgaatttatacactcttccaaatata	oIT146
Recombineering Oligo used in the stop codon insertion into *llnz_01155*	c*g*c*c*t*caactacatcactaccatcattttcagctttagcttttatgggttatcatgtatcttcgctatccaagtatgttactataaagtc	oIT147
Recombineering Oligo used in the stop codon insertion into *llnz_01160*	t*t*t*a*t*ttgtctgaatatggtgtgagaataacgatttatagggttatcaatccttaaatatagtgtctgagtctattttagctttggcccc	oIT148
Recombineering Oligo used in the stop codon insertion into *llnz_01175*	g*a*a*g*t*aactttgattgataaattcgtagtcaaagttttatccgcgatatcaagcacttacagtatgtccttcatagaaagtaaagagtaa	oIT149
Recombineering Oligo used in the stop codon insertion into *llnz_01175*	a*c*a*a*a*atgataaacaccatctggttcgttaaagaacggcatgggttatcaggataggataattccaaagaaaagagcaattgacaaataa	oIT150

Footnote**:** *=5' phosphorothioate nucleotide modifications

4.3.3 Primer extension analysis and Reverse Transcription PCR

L. lactis NZ9000 harbouring individual pPTPL constructs that exhibited promoter activity were subcultured in GM17 to an OD_{600nm} of ~0.1 and were allowed to grow at 30 ^{o}C until the culture reached mid-exponential growth phase, after which total RNA was extracted as described above. Primer extension was performed as previously described (22). Sequence ladders of the presumed promoter regions immediately upstream of genes with locus tags *llnz_01075*, *llnz_01135*, *llnz_01160*, and *llnz_01175* (of *L. lactis* NZ9000) were obtained from their corresponding pPTPL-based plasmid preparations. These sequence ladders were produced using primers (Table 2) that were also employed for the corresponding primer extension reaction using the Thermo Sequenase primer cycle sequencing kit according to manufacturer's instructions (Amersham). Separation of extension and sequencing products was achieved on a

10 % Li-Cor Matrix KB Plus acrylamide gel. Signal detection and image capture were performed in a Li-Cor sequencing instrument (Li-Cor Biosciences).

The transcriptional organization of the *cwps* gene cluster was confirmed using reverse transcription PCR (RT-PCR). Four independently grown cultures of *L. lactis* NZ9000 (20 ml) had their total RNA isolated (at OD_{600nm} of ~0.5) using a previously described method (23) in combination with High Pure RNA Isolation Kit (Roche Diagnostics, Mannheim, Germany). The RNA was then reverse-transcribed into cDNA using SuperScript™ III Reverse Transcriptase (Invitrogen, USA) along with a 9 bp long random-nucleotide primer, according to the manufacturer's instructions. The final cDNA preparation was used as a template to confirm the transcriptional separation (or connection by means of transcriptional read-through) of particular *cwps* genes by means of a PCR reaction targeting the regions directly upstream and downstream of the identified *cwps* promoters (Table 2). Regions within each deduced transcriptional unit were subjected to cDNA-based PCR amplification to serve as positive controls, while cDNA corresponding to regions upstream and downstream of the complete gene cluster were targeted by PCR as negative controls (Table 2).

4.3.4 CRISPR-Recombineering

CRISPR-cas9-assisted recombineering was adapted from a previously published method (16) and was applied to obtain mutations in 15 out of the 22 identified genes of the *cwps* gene cluster of *L. lactis* NZ9000, while two mutants were obtained using the 'classical' recombineering method. Recombineering was not attempted for genes that are proposed to be involved in the dTDP-l-rhamnose biosynthetic pathway (*rmlA, rmlB, rmlC,* and *rmlD*) as they are believed to be essential for bacterial growth. Briefly, a novel plasmid named pCNR was constructed, which contains the replication genes *repA, repD, repE* from the backbone of the pPTPi plasmid (17) combined with the chloramphenicol resistance gene (cm^r), P^{nisA} promoter and the ss-DNA binding protein-encoding gene, *recT*, from plasmid pJP005 (24) (see Chapter 2 of this thesis).

Subsequently, the single-step approach to CRISPR-Cas9-assisted recombineering was employed as previously outlined (17). *L. lactis* NZ9000 cells carrying both pVPL3004 (expressing Cas9 and tracrRNA) and pCNR (RecT-expressing plasmid) were made competent as previously described (24). These were then co-transformed with both 100 μg of the recombineering oligonucleotide (Table 3) and 100 ng of the corresponding pCRISPR construct. After recovery, cells were plated on GM17 agar plates supplemented with erythromycin (5 μg/ml) and tetracycline (10 μg/ml) selecting for plasmids pVPL3004 and pCRISPR, respectively. Colonies were screened using colony PCR and those containing the expected sequence insertion were further purified. The recombineering oligonucleotides were ordered from Integrated DNA Technologies (Leuven, Belgium).

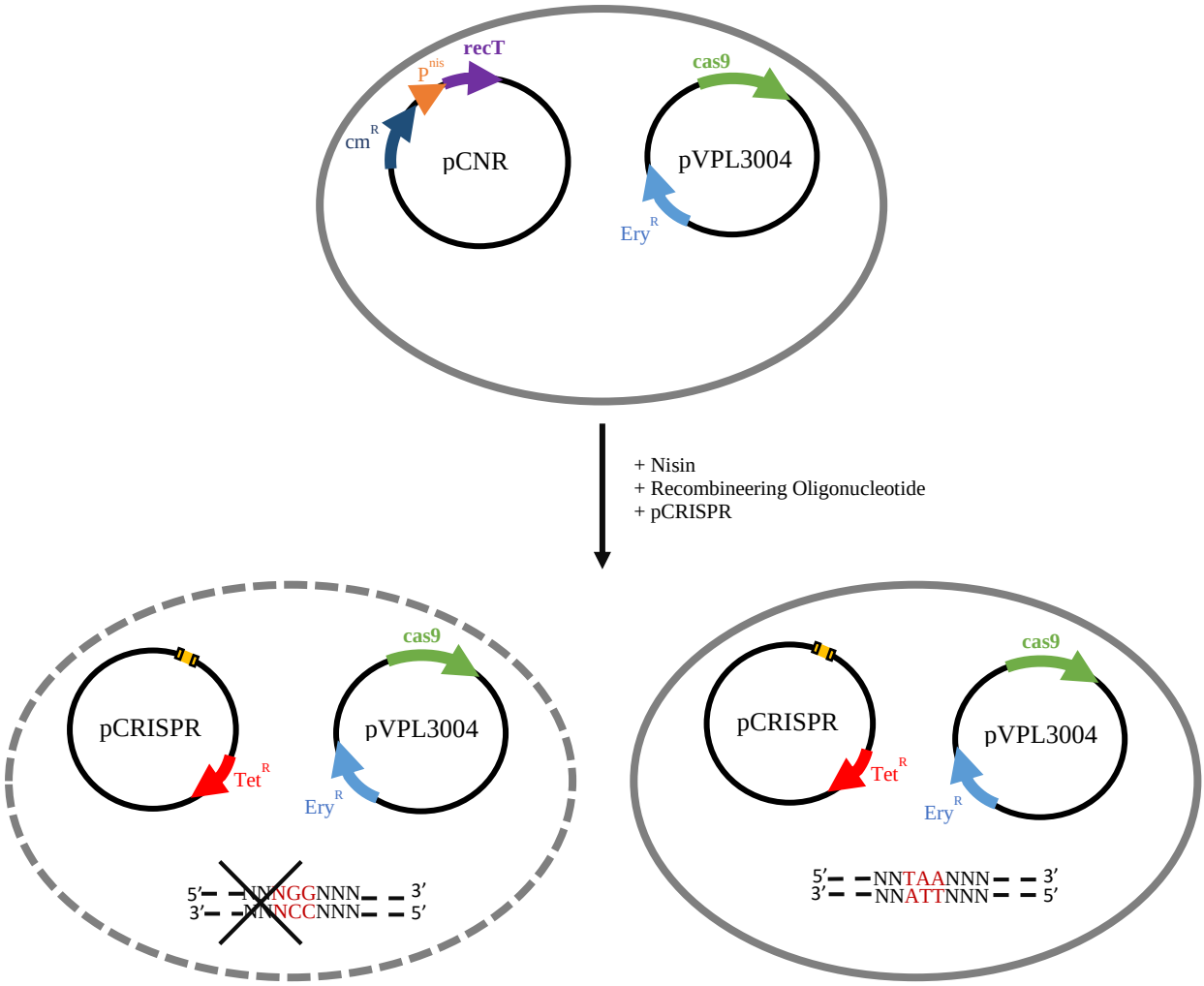


Figure 1. Schematic representation of the CRISPR-Cas9-assisted recombineering employed in this study. *L. lactis* NZ9000 carrying pCNR and pVPL3004 is made electrocompetent and the *recT* gene in pCNR is induced by the addition of nisin. The recombineering oligonucleotide carrying a non-sense mutation in the gene of interest that also removes the NGG protospacer adjacent motif needed for Cas9 recognition and cleavage. Finally, the pCRISPR carrying the targeting sequence matching that of the non-sense mutation is transformed into the *L. lactis* strain. Any strains that successfully incorporate the recombineering oligonucleotide into their genomes are expected to bypass the CRISPR-Cas9 nuclease, whereas strains retaining the wild-type sequence will be eliminated from the total population through the action of Cas9.

4.3.5 CWPS structural determination

Bacteria were harvested from an exponentially growing culture at an OD_{600} of 0.3 and cell walls were prepared as described previously (25). Briefly, heat-killed bacteria were boiled in 5 % SDS and the pellet, recovered by centrifugation, was treated successively by pronase and trypsin to remove proteins, and RNase and DNase to remove nucleic acids. The resulting material corresponding to purified cell walls (containing peptidoglycan and CWPS) was treated with 48 % HF for 48 h at 4 °C, to extract CWPS. The samples were centrifuged at 20,000 x *g* and the supernatant containing CWPS was dried under a stream of nitrogen. The residue was solubilized in Milli-Q H_2O and lyophilized. Rhamnan and PSP oligosaccharides present in the sample were separated by SEC-HPLC as described previously (11). Elution was performed with Milli-Q H_2O and detection of eluted compounds was performed with a refractometer (2414 Refractive Index Detector, Waters) and/or UV detector at 206 nm. Fractions corresponding to peaks containing rhamnan and PSP oligosaccharides were collected and dried under vacuum. They were further analyzed by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) using 2,5-dihydroxy-benzoic acid (DHB) matrix with an UltrafleXtreme instrument (Bruker Daltonics) (localized on the MetaboHUB platform, Institut Joliot, CEA, Université Paris Saclay, France).

4.3.5 Determination of fitness of *cwps* mutants

The *cwps* gene mutant strains obtained by recombineering or CRISPR-Cas assisted recombineering (Table 1) were evaluated to compare their growth profiles with that of the parental strain *L. lactis* NZ9000. Three independent cultures of a given strain were grown overnight in M17 supplemented with 0.5 % w/v glucose and antibiotics, where relevant, for the maintenance of the CRISPR-Cas assisted recombineering plasmids. The following day a 5 % subculture of the triplicate cultures was performed in the same media and growth was assessed for 8 hr by means of optical density measurements at 30 minute intervals.

4.3.6 Bacteriophage isolation and assays

Phages used in this study are listed in Table 1. Propagation of phages on their respective host strains was performed as previously described (26). Similarly, both spot/plaque assays (27) and adsorption assays (28) were performed as previously described. Phage escape mutants (Table 1) derived from the 936 group phage sk1 (NC_001835.1) and capable of infecting the PSP-deficient derivative of *L. lactis* MG1363 (VES5751) (8) were also isolated. Briefly, a single plaque isolate of sk1 was propagated in 10 ml culture of *L. lactis* NZ9000 and was filtered through a 0.45 μm filter, resulting in a lysate of approximately 10^8-10^9 pfu/ml. This lysate was then used to perform a large scale (4 L) propagation of the phage on *L. lactis* NZ9000 and the resulting lysate was then concentrated to 4 ml of lysate using polyethylene glycol 8000 (PEG_{8000}) precipitation as described previously (29). Finally, 100 μl to 1 ml of the concentrated phage lysate (10^{11}-10^{12} pfu/ml) was employed to challenge *L. lactis* VES5751, after which any visible plaques, representing potential escape mutants, were propagated on *L. lactis* VES5751 (See Chapter 5 of this thesis).

4.4 Results

4.4.1 Promoter mapping of the *cwps* gene cluster

The *L. lactis* NZ9000 *cwps* gene cluster contains a total of 22 ORFs (the locus tags and associated gene names are presented in Table 4). The predicted functions of the enzymes encoded by these genes, based on available annotation data and HHPred analysis of their amino acid sequence, as well as the total number of predicted transmembrane regions are summarized in Table 4. In order to identify potential promoter-containing regions within the *cwps* region of *L. lactis* NZ9000, any intergenic region present within the *cwps* cluster was amplified and cloned upstream of the *lacZ* gene located on the pPTPL vector (Figure 2). Promoter activity of the inserted DNA fragments in the generated pPTPL constructs was first assessed visually by blue/white colony assessment on GM17 agar plates containing X-gal (data not shown). Of the eight intergenic regions tested from the *cwps* gene cluster of *L. lactis* NZ9000, four (NZProm1, NZProm5, NZProm7, and NZProm8) appeared to contain an active promoter as their insertion into pPTPL corresponded to the generation of blue colonies (Figure 2). The relative strength of each of the potential promoter-containing regions identified above was measured as specific β-galactosidase activity and assessed at three different time points during growth. The specific β-galactosidase activities produced by *L. lactis* NZ9000::*NZProm1*, NZ9000::*NZProm5*, NZ9000::*NZProm7*, NZ9000::*NZProm8* at different time points during growth are presented in Figure 3. From the obtained results it is clear that the expression levels supported by the promoter sequences present in each of the intergenic regions are distinctly different, with NZ9000::NZProm1 consistently providing the highest level of expression (~800 - 900 Miller units). The remaining three intergenic regions exhibit progressively less activity with the last two, NZProm7 and NZProm8, exhibiting barely measurable β-galactosidase activity (~5.5 – 24 Miller Units).

Table 4. Proteins encoded in the *L. lactis* NZ9000 *cwps* locus and their predicted functions

Locus tag		Gene name	TM[1]	BlastP Predicted Function	HHPred Predicted Function [e-value, Top Hit]
NZ9000	MG1363				
llnz_1075	*llmg_0206*	*rmlA*	0	Glucose-1-phosphate thymidylyltransferase	Glucose-1-phosphate thymidylyltransferase (EC 2.7.7.24); [3.1^{-41}, 1MC3_B]
llnz_1080	*llmg_0207*	*rmlC*	0	dTDP-4-dehydrorhamnose 3,5-epimerase	dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase (E.C.5.1.3.13); [1.2^{-41}, 1NXM_A]
llnz_1085	*llmg_0208*	*rmlX*	0	Hypothetical protein	Hypothetical protein
llnz_1090	*llmg_0209*	*rmlB*	0	dTDP-glucose 4,6-dehydratase	dTDP-glucose 4,6-dehydratase (E.C.4.2.1.46); [3.0^{-45}, 1OC2_A]
llnz_1095	*llmg_0210*	*rmlD*	0	dTDP-4-dehydrorhamnose reductase	dTDP-4-dehydrorhamnose reductase; [5.0^{-37}, 4WPG_A]
llnz_1100	*llmg_0211*	*rgpA*	0	Alpha-D-GlcNAc alpha-1,2-L- rhamnosyltransferase	N-acetyl-alpha-D-glucosaminyl L-malate synthase (E.C.2.4.1.-); [8.9^{-37}, 5D01_A]
llnz_1105	*llmg_0212*	*rgpB*	0	alpha-L-Rha alpha-1,3-L-rhamnosyltransferase	Putative glycosyltransferase protein; [8.0^{-26}, 3BCV_A]
llnz_1110	*llmg_0213*	*rgpC*	6	ABC transporter permease	ABC transporter, Transport permease protein; [5.1^{-27}, 6AN7_D]
llnz_1115	*llmg_0214*	*rgpD*	0	ABC transporter ATP-binding protein	ABC-type dipeptide/oligopeptide/nickel transport system; [9.6^{-40}, 4FWI_B]
llnz_1120	*llmg_0215*	*wpsJ*	9	Membrane protein	4-amino-4-deoxy-L-arabinose (L-Ara4N) transferase; [4.3^{-9}, 5EZM_A]
llnz_1125	*llmg_0216*	*rgpE*	0	Glycosyltransferase family 2 protein	Putative glycosyltransferase (GalT1); [7.4^{-27}, 5HEA_C]
llnz_1130	*llmg_0217*	*rgpF*	0	Alpha-1,2-L-rhamnosyltransferase	Putative glycosyltransferase protein; [0.94, 3BCV_A]
llnz_1135	*llmg_0218*	*wpsA*	0	Glycosyl transferase family 2 protein	Dolichol monophosphate mannose synthase; [1.5^{-28}, 5MLZ_A]
llnz_1140	*llmg_0219*	*wpsB*	3	Hypothetical protein	Hypothetical protein
llnz_1145	*llmg_0220*	*wpsC*	0	Glycosyl transferase family 2 protein	Putative glycosyltransferase protein; [1.5^{-29}, 3BCV_A]
llnz_1150	*llmg_0221*	*wpsD*	0	Glycosyltransferase family 2 protein	Putative glycosyltransferase (GalT1); [1.0^{-28}, 5HEA_C]
llnz_1155	*llmg_0222*	*wpsE*	0	Glycosyltransferase	Hypothetical protein
llnz_1160	*llmg_0223*	*wpsF*	0	Galactofuranose transferase	Glycosyltransferase C; [1.0^{-30}, 4W6Q_A]
llnz_1165	*llmg_0224*	*glf1*	0	UDP-galactopyranose mutase	UDP-galactopyranose mutase (E.C.5.4.99.9); [2.1^{-69}, 4MO2_B]
llnz_1170	*llmg_0225*	*wpsG*	13	Flippase	Putative lipid II flippase MurJ; [5.7^{-36}, 5T77_A]
llnz_1175	*llmg_0226*	*wpsH*	2	Hypothetical protein	Hypothetical protein
llnz_1180	*llmg_0227/ 228*	*wpsI*	9	Hypothetical protein	4-amino-4-deoxy-L-arabinose (L-Ara4N) transferase; [1.4^{-9}, 5EZM_A]

Footnote: [1]TM = Transmembrane regions, as predicted by TMHMM Server v. 2.0

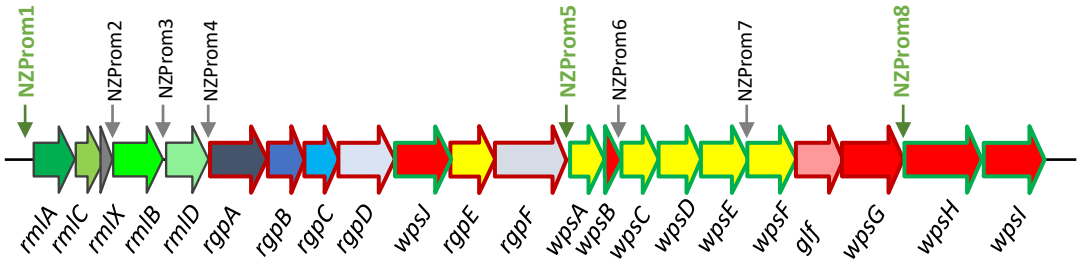


Figure 2. Schematic representation of the *L. lactis* NZ9000 cwps gene cluster encoding rhamnan and PSP biosynthesis. The intergenic regions that were cloned into pPTPL vector for promoter probing are also highlighted over the operon (*NZProm1-8*) (Table 2). Intergenic regions highlighted with a green arrow contain a putative promoter based on X-gal blue/white colony assays. Genes with successfully introduced non-sense mutations are highlighted with a green border while genes for which non-sense mutations could not be obtained are highlighted by a dark-red border.

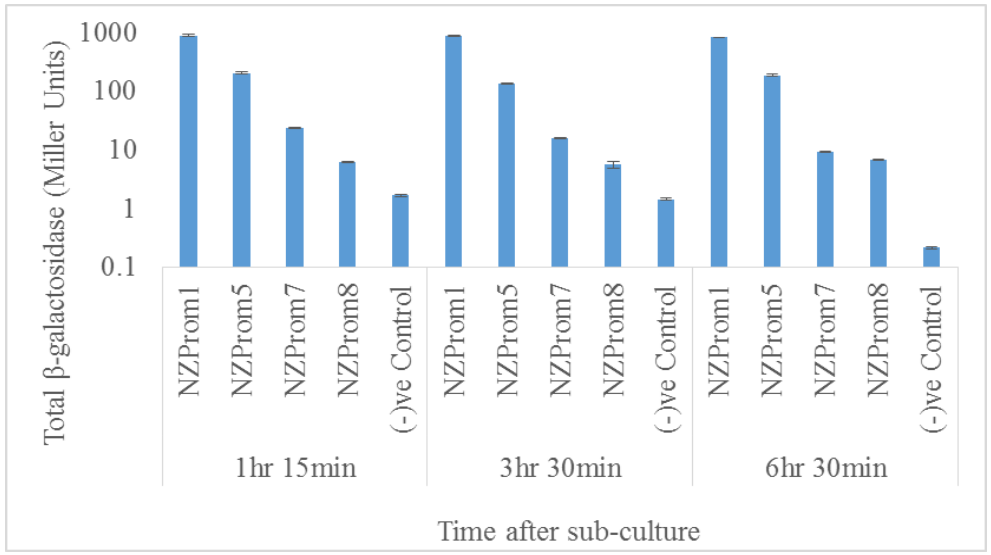
4.4.2 Transcription start site identification and RT-PCR

Following identification of potential promoter-containing regions within the *cwps* gene cluster of *L. lactis* NZ9000, the presence of promoter sequences was validated by means of primer-extension analysis. The four promoter-containing regions identified within *L. lactis* NZ9000 *cwps* gene cluster are upstream of the *rmlA* (corresponding to fragment *NZProm1*), *wpsA* (*NZProm5*), *wpsF* (*NZProm7*) and *wpsH* (*NZProm8*) (Figure 2). The gene with locus tag *llnz_01075* is predicted to encode a glucose-1-phosphate thymidylyltransferase (RmlA), the first of four genes encoding enzymes involved in the biosynthesis of dTDP-L-rhamnose, all of which are essential for growth (30-33). A transcription start site (TSS) was identified 28 nucleotides upstream of the annotated translation start site of *rmlA* (Figure 3B). Gene *wpsA* is predicted to encode a glycosyltransferase belonging to the GT-A family, and exhibits 43 % amino acid identity to GacI of *Streptococcus pyogenes,* which is reported to be involved in the decoration of the polyrhamnose cell wall polysaccharide with GlcNAc moieties (34). A TSS was identified 60 nucleotides upstream of the annotated translation start site of *wpsA* (Figure 3B). Gene *wpsH* is predicted to encode a membrane-associated protein and an extension product was identified 94 nucleotides upstream of the annotated translation start site of *wpsH*

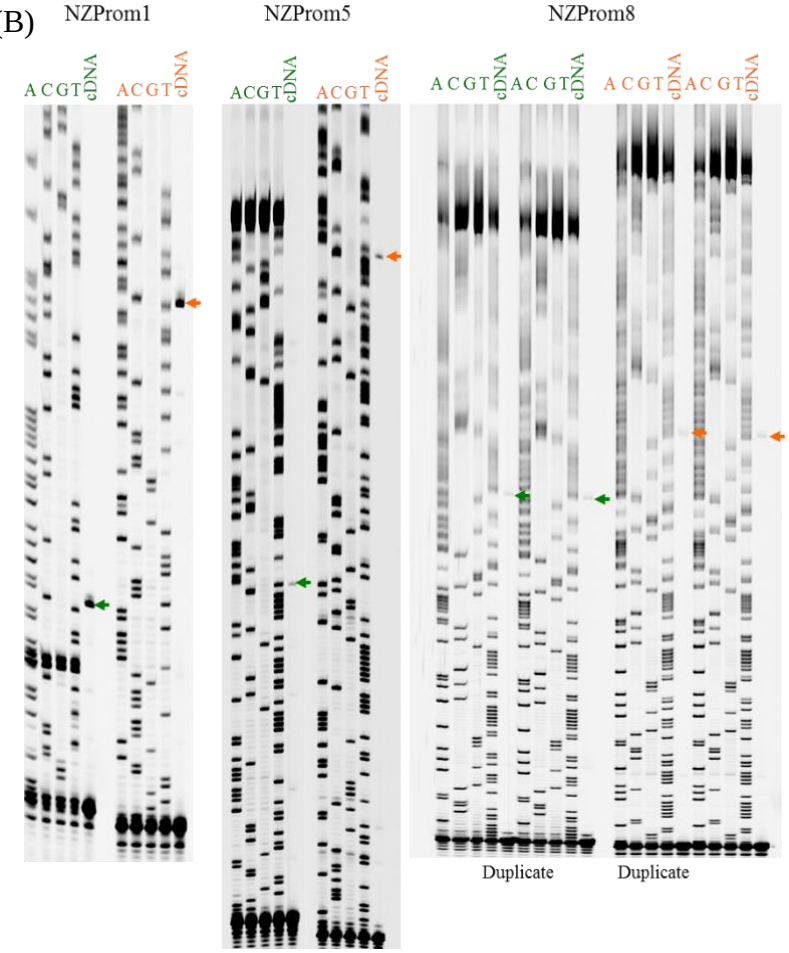
(Figure 3B). Finally, despite repeated attempts to identify an extension product for the promoter-containing region of *wpsF* no conclusive TSS was identified, which indicates that this region does not contain an active promoter but that perhaps a promoter was instead serendipidously created during the cloning of the intergenic region into the pPTPL promoter. Based on the TSS for *rmlA*¸ *wpsA*, and *wpsH* the -10 and -35 hexamers were manually identified (Figure 3C). The presumed -10 promoter recognition sequences of the *L. lactis* NZ9000 *cwps* gene cluster closely resemble the conserved motif sequence of the -10 regions in the MG1363 transcriptome, whereas the deduced -35 hexamers (as based on an approximate distance of 17 base pairs upstream of the -10 sequence) are not particularly similar to the consensus -35 motif (35). In addition, a dinucleotide TG motif was identified upstream of all identified -10 hexamers in the *cwps* gene cluster promoter regions. Such a motif has previously been shown to enhance promoter activity and is commonly found upstream of the -10 motif in lactic acid bacteria (36).

Despite the presence of at least three promoter-containing regions (*NZProm1*, *NZProm5, NZProm8*) and the successufully identified TSS for each of them, our analysis through RT-PCR hinted at the occurrence of transcriptional read-through between the two potential transcriptional subunits of the gene cluster. In more detail, through PCR amplification of the *L. lactis* NZ9000 cDNA, we were able to amplify regions directly overlapping both *NZProm5* and *NZProm8,* while the results for both the positive and negative control amplifications were according to expectations (Figure 4A). In conclusion, through the promoter mapping, primer extension, and RT-PCR analysis, we are proposing the presence of three transcriptional units (TU) within the *cwps* gene cluster: TU_{rha} under the control of *NZProm1*, TU_{psp} under the control of *NZProm5*, and TU_{pol} under the control of *NZProm8* (Figure 4B).

(A)



(B)



(C)

Strain	Gene / Locus Tag	-35 TTGACA	-10 TATAAT	+1 TSS
NZ9000	*rmlA/llnz_01075*	ATAGAAGG**TTGTGT**ATTGAGAAAGTTTTTG**TG**T**TATAAT**GTATATT**G**		
	pspA/llnz_01135	TTGAAAATAG**GCAAAA**AAAATAAGTAAATTTG**TG**G**TAAAAT**TTATATT**A**		
	pspH/llnz_01175	TGCATAATC**TTATAT**TTCAAGTATTGATTGG**TG**C**TATAAT**ATATATA**A**		

Figure 3. (A) Specific β-galactosidase activity of the strains *L. lactis* NZ9000 pPTPL::*NZProm1*, ::*NZProm5*, ::*NZProm7*, ::*NZProm8* and *L. lactis* NZ9000 pPTPL measured during the lag, mid-exponential, and late-exponential phase of growth. (B) 10 % Li-Cor Matrix KB Plus acrylamide gel including the sequence ladders and primer extension reactions for the promoter containing regions *NZProm1*, *NZProm5*, and *NZProm8*. (C) Summary of the identified promoter-containing regions along with the transcription start sites for each gene. Each of the -10 and -35 hexamer sites was manually identified based on nucleotide distance from the TSS. The conserved -10 and -35 hexamers of MG1363 as were previously identified are also indicated [30]. Highlighted in green is the dinucleotide TG motif found commonly upstream of lactic acid bacteria -10 promoter region [31].

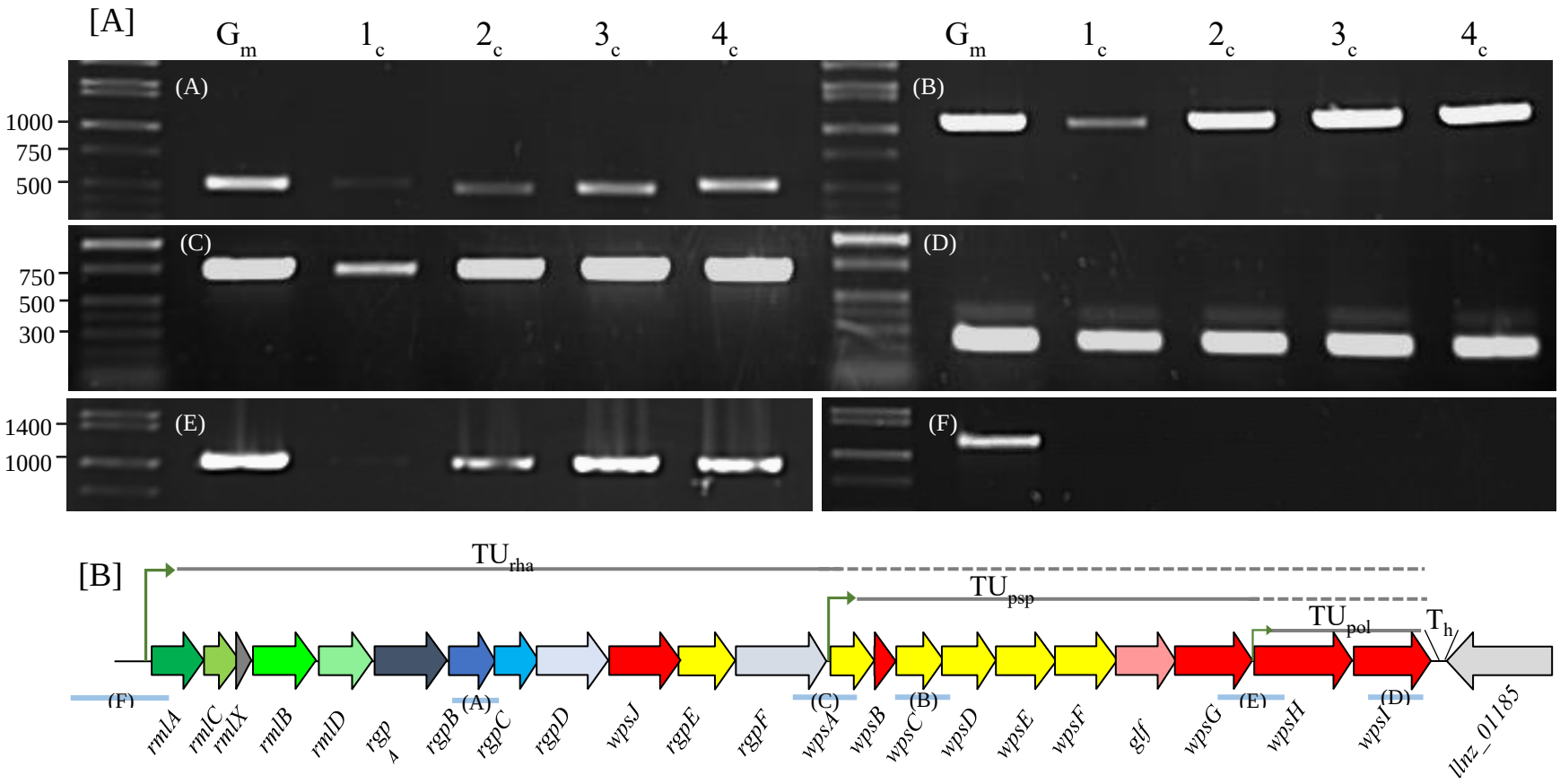


Figure 4. [A] Gel electrophoresis summarizing the RT-PCR results for the transcriptional organization of the *cwps* gene cluster. Following total mRNA isolation at mid-exponential growth from four biological independent cultures of *L. lactis* NZ9000 (1_c-4_c) and conversion into cDNA, the transcriptional organization of the gene cluster was examined through PCR amplification of five regions within the cluster as well as one flanking *NZProm1* region as negative control. Genomic DNA (G_m) was also included for every set of primers as a positive control. PCR amplifying regions within or directly flanking (A) *rgpB*, (B) *wpsC*, (C) *NZProm5*, (D) *wpsI*, (E) *NZProm8*, (F) *NZProm1*. Numbers on the left-hand side of the figure indicate the total size of the amplified DNA in base pairs. [B] Schematic representation of the *L. lactis* NZ9000 *cwps* gene cluster highlighting its deduced transcriptional organization. The three identified promoter regions are shown by green arrows ahead of *rmlA*, *wpsA*, and *wpsH*. Three transcriptional units based on these promoter regions are also identified with grey lines (dashed lines = potential transcriptional read-through identified by the RT-PCR above). The regions amplified during RT-PCR are also highlighted with light blue bands below the cluster. Finally, the end of the gene cluster is highlighted by the inclusion of the first ORF (*llnz_01185*) downstream of it as well as the deduced terminator hairpin (T_h) as deduced by the online platform ARNold. TU_{rha} = Rhamnan transcriptional unit; TU_{psp} = PSP transcriptional unit; TU_{pol} = Polymerization module transcriptional unit.

4.4.3 *L. lactis* NZ9000 *cwps* cluster non-sense mutations and CWPS structural analysis

In order to perform a functional analysis of individual genes of the *L. lactis* NZ9000 *cwps* gene cluster, nine non-sense mutants (Table 1) were created by CRISPR-recombineering. With the exception of the first five genes (*rmlABXCD*) in the cluster that are believed to be part of the *rml* pathway for the production of the precursor molecule for UDP-rhamnose, mutational attempts were made to target all other genes of the *cwps* cluster. Notably, we were unable to obtain eight out of the 17 attempted non-sense mutants. The majority of these eight genes, which could not be successfully mutated, form part of the rhamnan assembly pathway as it has been previously characterized (11). Attempts by other groups to knock-out these genes through double-crossover homologous recombination were similary unsuccessful and this appears to indicate that this set of genes is essential for the survival and growth of the strain (11). In total, nine non-sense mutants were, however, successfully isolated (designated *L. lactis* NZ9000-*wpsA, -wpsB, -wpsC, -wpsD, -wpsE, -wpsF, -wpsH, -wpsI, -wpsJ;* Table 1 and Figure 2). R

The CWPS of each of the successfully obtained mutants was extracted from isolated cell walls by HF treatment and separated by SEC-HPLC, as described previously (11). For the control wild-type *L. lactis* NZ9000, HF treatment allows extraction of rhamnan from the cell wall and generating oligosaccharides by cleavage of the PSP chains at the phosphodiester bonds. The two generated major polysaccharide-containing peaks, corresponding to rhamnan and PSP oligosaccharides, were then further separated and analysed by SEC-HPLC (Figure 5). This analysis of the CWPS extracted from the nine obtained mutants revealed first that all mutants were able to synthesize rhamnan (a subset of these nine mutants are shown in Figure 5). Notably, inactivation of *wpsJ*, a gene located in the more conserved part of the *cwps* gene cluster (which are believed to be involved in rhamnan synthesis) did not prevent the synthesis of rhamnan. In contrast, PSP synthesis was profoundly affected in eight out of the nine mutants,

with no or extremely low amounts of PSP extracted (Figure 5). Only mutant NZ9000-*wpsB* and NZ9000-*wpsI* continue to synthesize a substantially reduced amount of PSP when compared to WT NZ9000 (Figure 5).

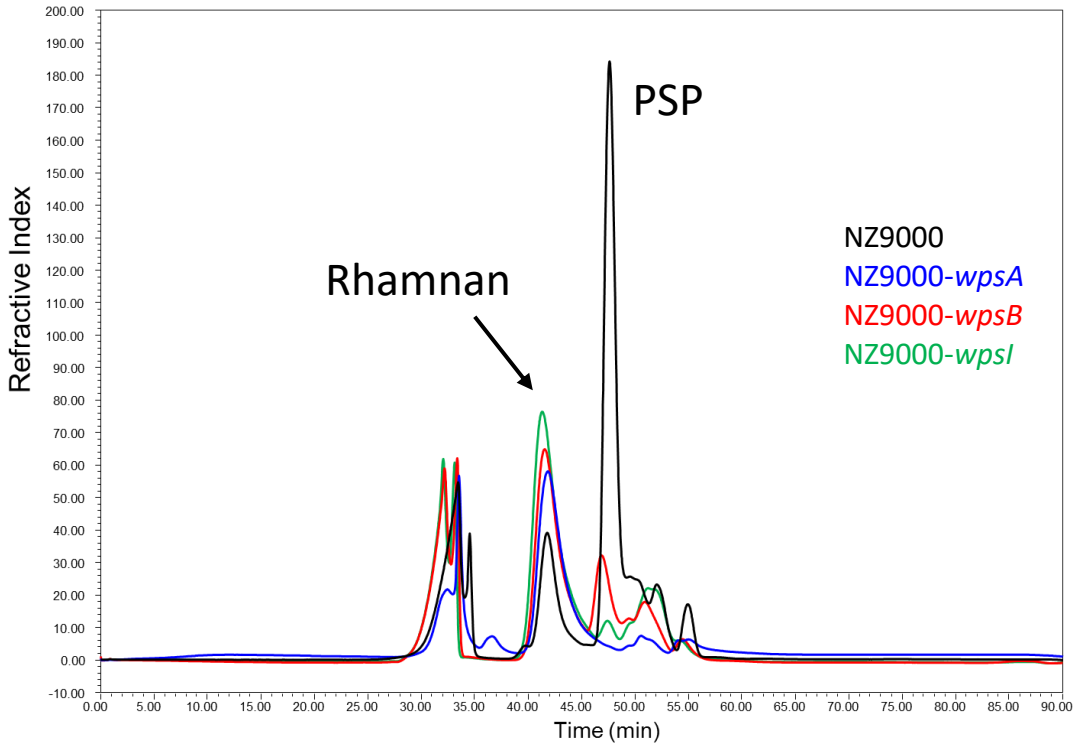


Figure 5. SEC-HPLC analyses of CWPS extracted from *L. lactis* NZ9000 and three *psp* mutants

Despite the apparent absence or very low amount of PSP in the mutants according to the SEC-HPLC profile, we reasoned that such mutants may still be able to synthesize small amounts of oligosaccharide originating from aborted synthesis of the PSP subunit because of the mutation. Thus, we employed the highly sensitive Maldi-Tof MS analysis to assess glycan content of the fractions with an elution time corresponding to wild-type PSP for each mutant. The spectra presented in Figure 6 show different patterns that correspond to: (i) detection of the hexasaccharide subunit similar to that of WT NZ9000 (mutant *wpsB*); (ii) detection of compounds with lower m/z values than would be expected for the complete hexasaccharide PSP subunit (for strains containing a mutation in *wpsD, wpsE, wpsF, wpsH* or *wpsI*); (iii) absence of any molecular species (for strains with a defunct *wpsA, wpsC* or *wpsJ*).

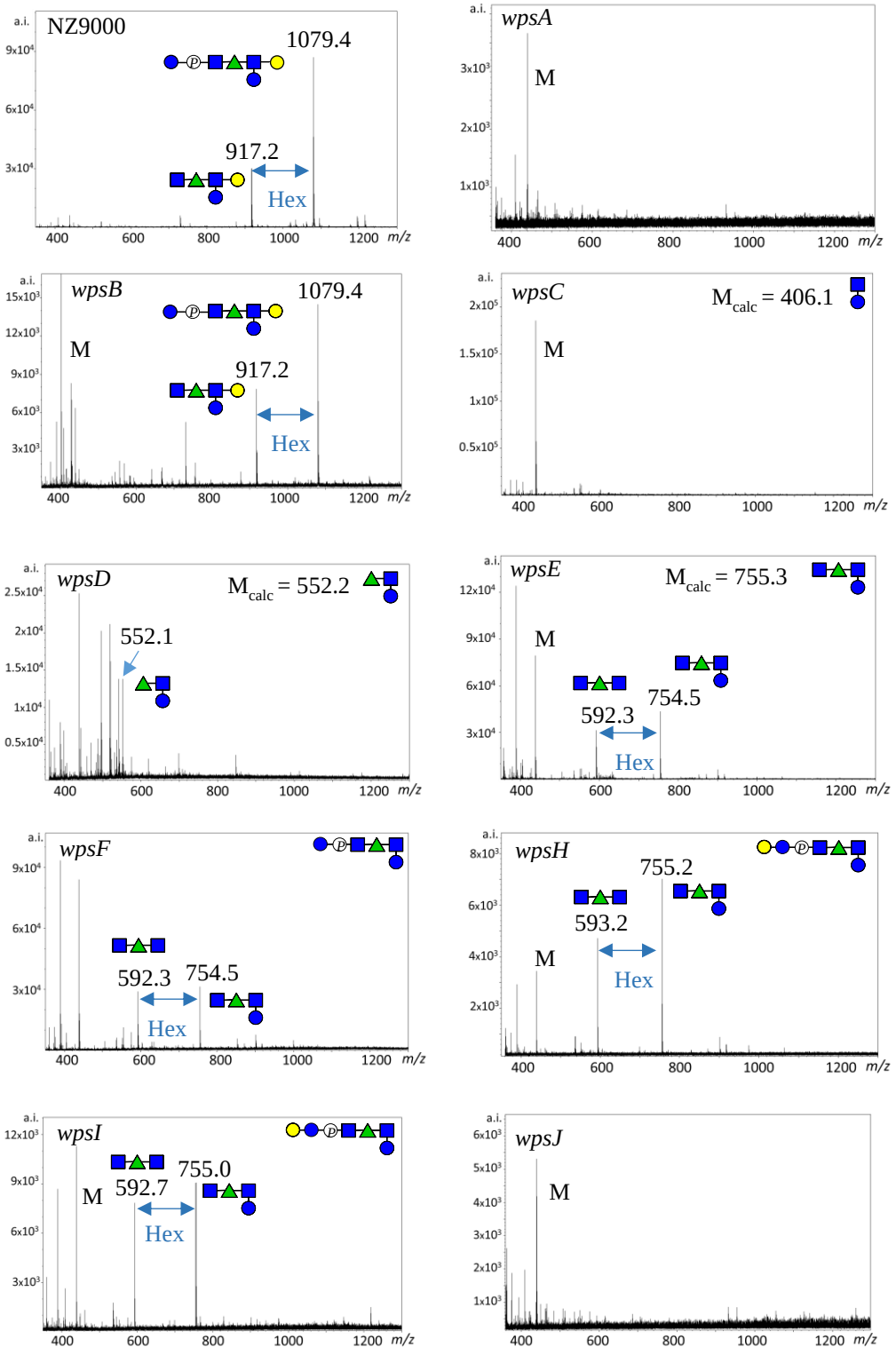


Figure 6. Maldi-Tof MS spectra of PSP fragments extracted by HF from the cell walls of *L. lactis* NZ9000 and various of its *psp* mutants. *m/z* values correspond to $[M+Na]^+$ adducts. *L. lactis* NZ9000 synthesizes polymeric PSP that is cleaved during HF extraction. A similar spectrum is found for *L. lactis* NZ9000-*wpsB*. For the remaining *psp* mutants, the proposed truncated PSP structure before HF cleavage of the phosphodiester bond is represented in the right top corner of the spectrum. Similarly, figures within the spectrum indicate proposed PSP structures based on *m/z* values. *m/z* values with a 162-m.u decrease are assigned to PSP fragments devoid of side-chain Glc. Hex = hexose; M = matrix; M_{calc} = calculated mass; blue box = GlcNAc; green triangle = Rha; blue circle = Glc; yellow circle = Gal*f*; *P* = phosphodiester bond.

4.4.4 Design of a tentative model for PSP biosynthesis and attachment to rhamnan

According to previous results and sequence analysis (9, 11, 12), the more conserved region of the *cwps* gene cluster of *L. lactis* appears to be responsible for rhamnan synthesis whereas the more variable region is involved in PSP biosynthesis. Transcriptional analysis of the *cwps* locus presented above also supports this hypothesis. In *L. lactis* MG1363/NZ9000, the variable region of the *cwps* gene cluster encodes five putative glycosyltransferases (*wpsA,C, D, E, F*), four membrane proteins (*wpsB, G, H, I*) and a gene (*glf*) encoding a protein with putative UDP-galactopyranoside mutase activity (Table 4). Based on protein sequence analysis of the encoded CWPS proteins (Table 4) and the results from CWPS analysis by MS as shown on Figure 6, we propose the following model for PSP biosynthesis and attachment to rhamnan.

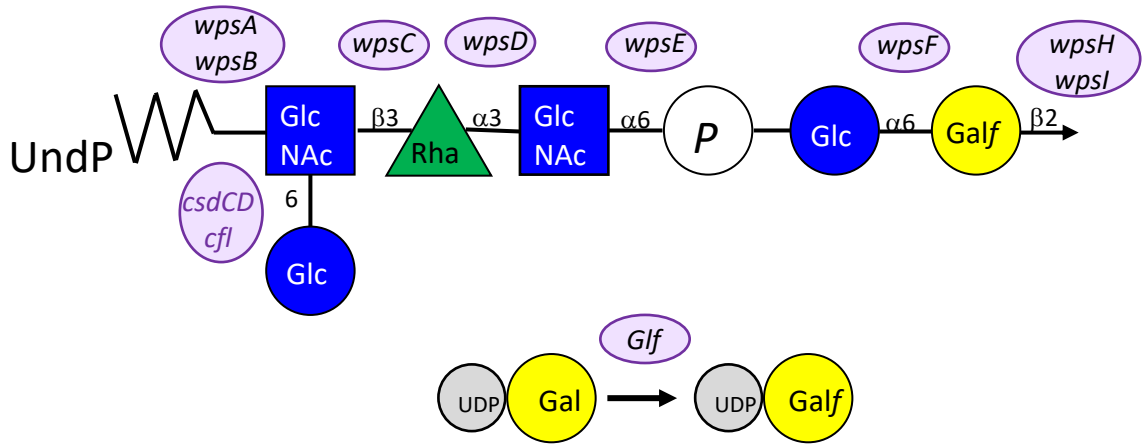


Figure 7. [Top part] Predicted structure of the UndP-linked PSP intermediate of *L. lactis* NZ9000 PSP and genes proposed to be involved in the synthesis of each bond. [Bottom part] Formation of UDP-Gal*f* for incorporation in the PSP repeat unit by Glf.

Initiation of PSP repeat unit synthesis- Owing to the presence of a predicted flippase-encoding gene (*wpsG*) in the variable part of the *cwps* gene cluster and because PSP is made of repeated phospho-hexasaccharide subunits, we propose that PSP is synthesized through a Wzx/Wzy-dependent pathway. Thus, initiation of the subunit synthesis should start in the cytoplasm on a lipid precursor, presumably Und-P (Figure 7). However, no typical priming transferase of polysaccharide biosynthesis of the WecA or WbaP families (37) is encoded within the variable part of the *cwps* cluster (Table 4). The two proteins encoded by the first two genes of the variable region (*wpsA* and *wpsB*) of the *cwps* cluster exhibit sequence homology with GacI and GacJ from *Streptococcus pyogenes* which were recently reported to be involved in the decoration of the polyrhamnose cell wall polysaccharide with GlcNAc moieties (38). GacI was shown by biochemical experiments to be a glycosyltranferase catalyzing the transfer of GlcNAc from UDP-GlcNAc onto Und-P to generate Und-P-GlcNAc, while GacJ stimulates GacI enzymatic activity.

Interestingly, our SEC-HPLC and MS analysis of CWPS (Figure 5 and Figure 6) indicates that no PSP-derived oligosaccharide was detectable in Peak 2 of mutant NZ9000-*wpsA*, whereas inactivation of gene *wpsB* allows the synthesis of PSP (albeit at a severely reduced amount when compared to WT). These results suggest that WpsA activity is absolutely required for PSP synthesis, whereas WpsB is not. According to its predicted enzymatic activity (GacI-like activity), we propose that WpsA catalyzes the first step of PSP subunit assembly and initiates synthesis by transferring GlcNAc from UDP-GlcNAc to Und-P and that WpsB (GacJ homolog) stimulates WpsA activity (Figure 7). This hypothesis is supported by the results of MS analysis, with no PSP fragments detected in NZ9000-*wpsA* mutant, but a spectrum similar to wild-type NZ9000 for NZ9000-*wpsB* (Figure 6). Of note, this model is unusual in that PSP repeat unit synthesis would take place on an Und-P-sugar precursor rather than the classical Und-P-P-sugar precursor (Figure 7).

Elongation of the PSP repeat unit – Downstream of *wpsA* and *wpsB* four genes are located that encode putative glycosyltranferases which we propose to be involved in the iterative synthesis of the PSP repeat unit as outlined on (Figure 7). First, glycosyltransferases WpsC and WpsD are proposed to elongate the PSP subunit on Und-P-GlcNAc by transferring Rha and GlcNAc from dTDP-Rha and UDP-GlcNAc nucleotide sugars, respectively. Then *WpsE*, which is endowed with a stealth domain found in phosphosugar transferases (Stealth_CR2; PF11380), would add Glc-phosphate. Finally, WpsF, being annotated as a galactofuranose transferase, would add galactofuranose (Gal*f*) using UDP-Gal*f*, which in turn is produced by Glf annotated as an UDP-galactopyranose mutase (Figure 7). Regarding the side-chain Glc, we have recently shown that the genetic determinants involved in the addition of this sugar on PSP subunits are located outside of the *cwps* gene cluster (See Chapters 2 and 3 of this thesis).

This model for the PSP biosynthesis pathway is corroborated by data obtained from the MS analyses. Regarding the two mutants NZ9000-*wpsD* and NZ9000-*wpsE*, molecular ions corresponding to oligosaccharide structures expected from the arrested synthesis of PSP subunit according to each mutation were detected in agreement with the proposed model (Figure 6 and Figure 7). Of note, deduced structures contain (at least partially) the side chain Glc. Regarding NZ9000-*wpsC*, the absence of any detected signal may be due to its low molecular weight.

Export and polymerization of the subunits- Once formed, the lipid-linked oligosaccharide would be flipped across the cytoplasmic membrane by WpsG, which possesses the characteristics of a flippase (Table 1). HHPred analysis of WpsG revealed significant structural similarity to lipid II flippase MurJ (probability 100 %, E-value 5.7^{-36}, Structure 5T77 in PDB database), which catalyzes the translocation of the lipid-linked precursor of the peptidoglycan across the lipid membrane. Notably, the corresponding gene could not be inactivated, perhaps due to an indirect effect on the availability of uncoupled Und-P that might negatively impact

the assembly of other essential cell wall components. Finally, the repeat units are expected to be polymerized on the outside of the cytoplasmic membrane. Two membrane proteins are encoded at the 3' end of the *cwps* locus. Neither of these proteins exhibit sequence similarity to known Wzy polymerases. From literature it is known that these polymerases, because they are specific for their respective substrates, do not possess particularly conserved sequences, yet typically represent integral membrane proteins containing 10-14 TM segments (39, 40). In our particular case, the PSP polymerase may possess specific features because it is dedicated to polymerize repeat units attached on an Und-P rather than an Und-P-P. WpsI is endowed with 10 TM segments, so we propose that this protein fulfills this role. Finally, WpsH with two TM segments could be the last component of the pathway, with a "co-polymerase" function and/or a role in controlling the number of repeats polymerized by WpsI, reminiscent of a function classically named Wzz in the Wzx/Wzy pathway (41). The MS spectra from the two mutants NZ9000-*wpsH* and *wpsI* also revealed the peak with m/z value of 755.2, being consistent with a single PSP subunit as represented in Figure 7, and in agreement with their proposed role in PSP polymerization.

Attachment of PSP to rhamnan - From our previous results (11), we gathered evidence that rhamnan is anchored to peptidoglycan, while PSP may be attached to rhamnan, thereby constituting a large heteropolysaccharide in the cell wall. The WpsJ protein which is an integral membrane protein with eight predicted TM segments does not exhibit sequence homology with any protein of known function. However, HHPred analysis revealed modest structural similarity to oligosaccharidyltranferase PglB (probability 91.19 %, E-value 0.49) structure 3WAJ in PDB database), which catalyzes the transfer of oligosaccharide chains to Asn residues in proteins. We hypothesize that this protein, being encoded by *wpsJ* located in the more conserved part of the *cwps* gene cluster, among the genes encoding rhamnan biosynthesis, is the transferase involved in attaching polymerized PSP to the rhamnan chain (Figure 7).

Regarding the NZ9000-*wpsJ* mutant, CWPS analysis by SEC-HPLC showed an absence of any PSP peak (profile similar to NZ9000-*wpsA*) and no signal could be detected by MS analysis of the fractions corresponding to the PSP peak (Figure 6). These results further support the involvement of WpsJ in PSP synthesis. They are in agreement with the proposed role of WpsJ in the transfer of PSP chains onto rhamnan, which would be a necessary step for PSP to be extracted following purification from cell walls as described in the Materials and Methods section.

4.4.5 Assessment of the model by phenotypic analysis of the *cwps* mutants.

The effect of individual non-sense mutations on growth was determined and compared to that of the control strain *L. lactis* NZ9000 (Table 1 and Figure 6). All nine *cwps* mutant derivatives presented with an impaired growth phenotype compared to the control strains both in liquid broth conditions, while they also formed smaller colonies on solid agar plates as compared to the wild-type strain (data not shown). In addition to the observed effect of the mutation on growth, the mutants also exhibit an increased level of sedimentation in liquid culture as compared to the control strains.

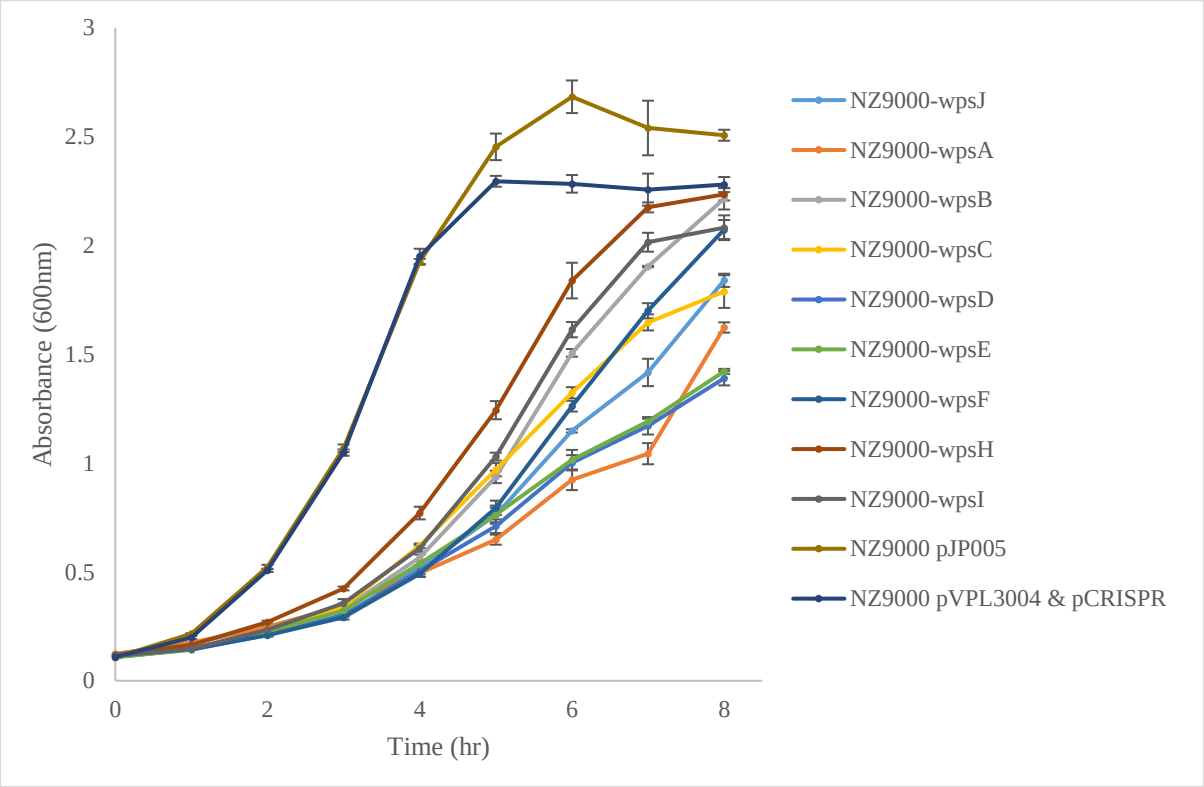


Figure 8. Growth curve of control strains *L. lactis* NZ9000::pJP005 and *L. lactis* NZ9000::pVPL3004 ::pCRISPR versus *L. lactis* NZ9000-*wpsJ*, -*wpsA*, -*wpsB*, -*wpsC*, -*wpsD*, -*wpsE*, -*wpsF*, -*wpsH*, -*wpsI*.

All *cwps* gene mutants were assayed for their sensitivity against phages that are able to infect the wild-type *L. lactis* NZ9000. As expected, all three phages assayed (jj50, p2, sk1) were unable to infect the majority CWPS mutant strains (with the exception of *L. lactis* NZ9000-*wpsB* and NZ9000-*wpsI*), presumably due to the PSP-deficient phenotype (Table 5). However, we were able to isolate derivatives of phage sk1 that showed infectivity against certain PSP-deficient strains. More specifically, eight escape mutants of sk1 (Table 1) were successfully isolated against *L. lactis* VES5751, another PSP-deficient strain (non-sense mutation in *wpsH*). Their infectivity profiles against all obtained CWPS-deficient mutants were then investigated. Interestingly, the escape mutants were able to infect only a subset of the CWPS mutants (Table 5); more specifically, the mutants whose structural analysis had indicated that remnants of the PSP were still present on the cell wall, albeit at significantly reduced levels. In contrast, CWPS mutants that were characterized as completely devoid of PSP (i.e. mutants NZ9000-*wpsJ*, -*wpsA*, -*wpsC*) or producing small amounts of trisaccharide (NZ9000-*wpsD*) remained completely resistant to the phage escape mutants. Finally, *L. lactis* NZ9000-*wpsE* retains resistance to three of the escape mutants while being infected by the remaining five (Table 5).

Table 5. Phage host range against *L. lactis* NZ9000 control strain as well as nine *cwps* gene mutant derivatives. The phage represented include three wild-type 936 lytic phages commonly infecting *L. lactis* NZ9000 as well eight escape mutant derivatives of phage sk1. Grey box = phage infection, White box = no infection.

		Bacteriophage										
		Classical			Evolved							
		jj50	p2	sk1	MCC1	MCC5	MCC17	IT1	IT2	IT3	IT4	IT5
Bacterial Strain	*L. lactis* NZ9000 pVPL3004 / pCRISPR	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	L. lactis NZ9000-*wpsJ*	×	×	×	×	×	×	×	×	×	×	×
	L. lactis NZ9000-*wpsA*	×	×	×	×	×	×	×	×	×	×	×
	L. lactis NZ9000-*wpsB*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	L. lactis NZ9000-*wpsC*	×	×	×	×	×	×	×	×	×	×	×
	L. lactis NZ9000-*wpsD*	×	×	×	×	×	×	×	×	×	×	×
	L. lactis NZ9000-*wpsE*	×	×	×	×	✓	✓	×	✓	×	✓	✓
	L. lactis NZ9000-*wpsF*	×	×	×	✓	✓	✓	✓	✓	✓	✓	✓
	L. lactis NZ9000-*wpsH*	×	×	×	✓	✓	✓	✓	✓	✓	✓	✓
	L. lactis NZ9000-*wpsI*	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

4.5 Discussion

The outermost part of the cell envelope of *Lactococcus lactis* MG1363, a CWPS type C strain (9, 12), has recently been shown to contain two distinct polysaccharides, which are believed to be covalently linked to each other. These form a thin compact outer layer formed by an acidic phosphopolysaccharide named the PSP or pellicle, as well as a neutral rhamnan polysaccharide that is embedded in and/or covalently attached to the peptidoglycan layer (11). Similar structural results for the cell wall of *L. lactis* 3107, another CWPS type C strain, have been observed (9, 11). The separation of PSP and rhamnan observed in CWPS type C strains, which is thought to be a result of the harsh extraction procedure and indicative of the linkage type between the two polysaccharides, was not observed in representative strains of either the CWPS type A (13) or type B (14). Instead, a linear rhamnan backbone seems to form the major component of the CWPS of strains containing these latter two CWPS types, along with regular or irregular polysaccharide substitutions, which appear to represent the equivalent of the PSP seen in CWPS type C strains. In the current study, we propose a model for PSP biosynthesis and attachment to rhamnan based on sequence analysis, mutant construction and CWPS analysis. This model is also supported by transcriptional data and phage sensitivity experiments. Furthermore, taking into account the previous model that has been designed for rhamnan synthesis (11), we present here a comprehensive model for the biosynthesis of the complete hetero-polysaccharide structure (Figure 9). This model also includes proteins encoded by genes located outside the *cwps* gene cluster, such as TagO shown to be involved in the initiation of rhamnan synthesis (11), LcpA proposed to be involved in rhamnan attachment to peptidoglycan (11), and proteins involved in Glc decoration on the rhamnan and PSP subunits (See Chapters 2 and 3 of this thesis).

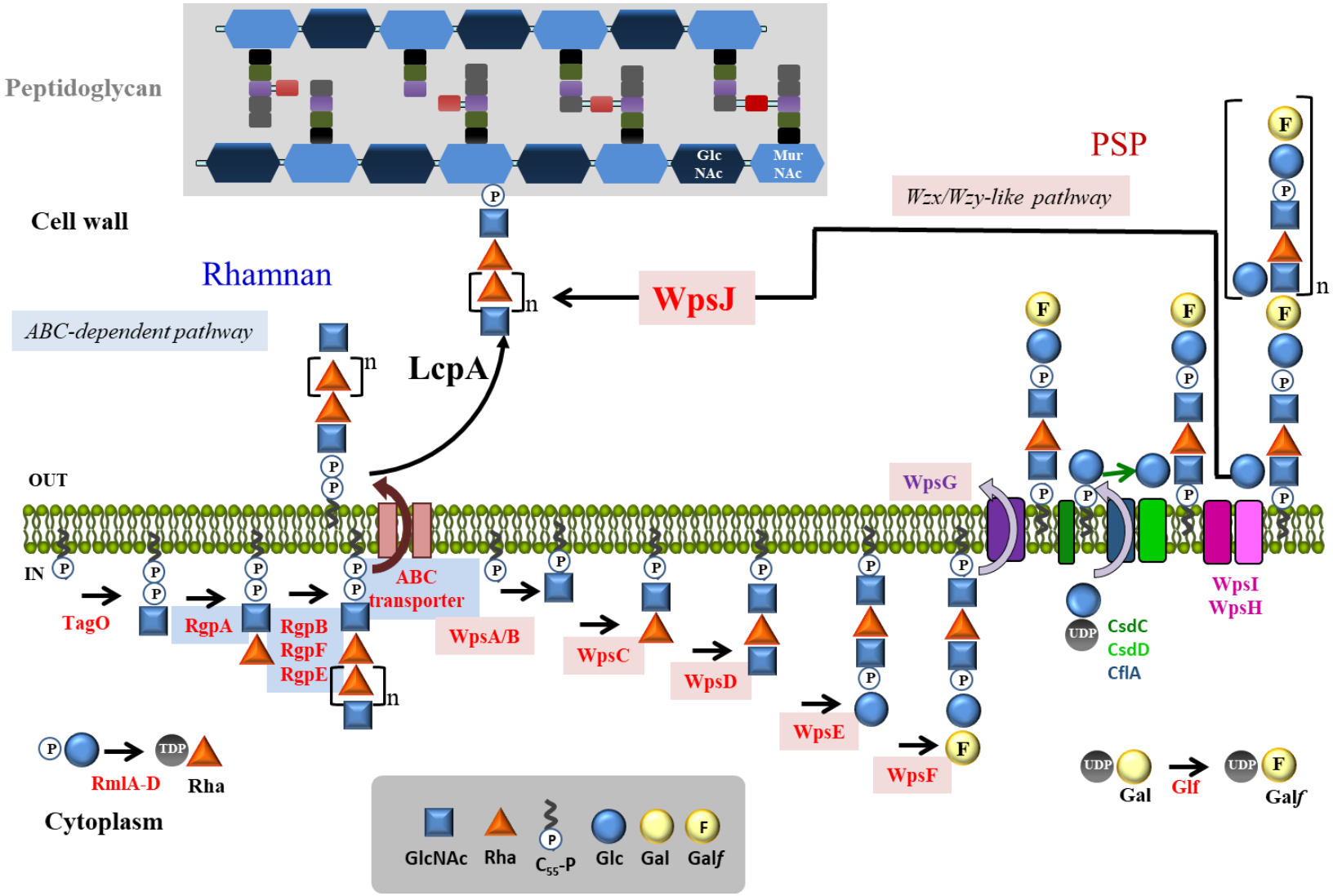


Figure 9. Schematic of the proposed enzymatic pathways involved in the assembly of rhamnan and PSP of *L. lactis* NZ9000

The proposed model constitutes a complex and original pathway, with the two glycan constituents, rhamnan and PSP, being synthesized independently by two distinct pathways, an ABC-transporter dependent pathway for rhamnan and a Wzx/Wzy-like pathway for PSP. In addition, we propose that the rhamnan and PSP subunits are synthesized as two different lipid-sugar intermediates. As shown in Figure 9, rhamnan subunit biosynthesis is initiated by TagO by transfer of GlcNAc-P on Und-P giving rise to Und-P-P-GlcNAc, whereas PSP repeat unit synthesis is initiated by WpsA and WpsB, belonging to the GacI/GacJ family, by transfer of GlcNAc onto Und-P giving rise to Und-P-GlcNAc. To our knowledge, this is the first model suggesting the synthesis of an oligosaccharide subunit on an Und-P-sugar precursor, before flipping and polymerization through a Wzx/Wzy-like pathway.

The conserved 5' part of the *cwps* contains first the *rmlABCD* genes encoding the four enzymes known to be involved in rhamnose precursor dTDP-L-Rha synthesis (42). After addition of GlcNAc-P on Und-P by TagO, the rhamnosyltransferase RgpA transfers the first rhamnosyl residue on the GlcNAc-P-P-Und lipid intermediate acceptor and initiates synthesis of the rhamnan chain. Elongation of the chain is performed by two other rhamnosyltransferases RgpB and RgpF. Polymerization until an average length of 33 Rha residues (corresponding to 11 repeating units) is terminated by addition of a GlcNAc or Glc residue at the non-reducing end of the chain, catalyzed by an as yet unknown enzyme (discussed further below). Together, RgpC and RgpD are proposed to catalyze transfer of the rhamnan chain to the outer side of the cytoplasmic membrane. Following export, the rhamnan is most probably anchored covalently to peptidoglycan acceptor sites, via phosphodiester linkage to MurNAc or GlcNAc residues of the glycan chains, with LcpA being the main ligase involved.

Regarding PSP, synthesis of GlcNAc-P-Und by WpsA/WpsB, the repeat unit is synthesized by the successive addition of Rha, GlcNAc, Glc-P and Gal*f* by the glycosyltranferases WpsC, WpsD, WpsE and WpsF, respectively, with UDP-Gal*f* formed from

UDP-Gal by Glf. The pentasaccharidic PSP repeat unit is flipped to the outside of the cytoplasmic membrane by WpsG. It is then polymerized by WpsI with WpsH as a putative co-polymerase and/or chain length regulator. Finally, WpsJ is involved in the transfer of the PSP chain on rhamnan. The site of linkage and the nature of the bond between the rhamnan and PSP constituents of CWPS remain to be determined. Also, the chronological order of rhamnan anchoring to peptidoglycan by an Lcp ligase and PSP branching on rhamnan by WpsJ is unknown. According to our unpublished data, the last sugar of the hexasaccharide repeat unit, the side chain Glc is added by a glycosylation system encoded by genes that are located outside the *cwps* cluster and we presume that this Glc is added outside the cytoplasmic membrane (See Chapters 2 & 3 of this thesis)

The last gene (annotated *rgpE*) without putative function in the *cwps* cluster could not be inactivated by any available genetic tool including the CRISPR approach. The *rgpE* gene encodes a glycosyltransferase with no specific characteristic, and it is present in the six strains of CWPS type A, B and C, for which sequence data is available (12). We propose that RgpE is involved in rhamnan chain termination, thus regulating rhamnan length by adding a sugar at the non-reducing end. Our previous MS analysis of rhamnan from *L. lactis* MG1363 and the absence of terminal Rha have indeed suggested the presence of a sugar (GlcNAc or Glc) at the non-reducing end of the rhamnan chain (11).

Und-P-monosaccharide lipid intermediates were previously described in different glycosylation mechanisms of cell wall glycopolymers such as LPS in Gram-negative bacteria or rhamnopolysaccharide, WTA and LTA in Gram-positive bacteria (38, 43). However, in these processes, the generated Und-P-monosaccharide is flipped outside the cytoplasmic membrane and the monosaccharide then becomes attached to a polysaccharidic chain. In our model, Und-P-GlcNAc serves as a precursor for synthesis of a pentasaccharidic linear repeat

unit and the Und-P-linked oligosaccharide is then flipped outside the cytoplasmic membrane followed by polymerization on this same Und-P precursor.

The three membrane proteins, WpsG, WpsH and WpsI, are proposed to play similar roles to Wzx/Wzy/Wzz to synthesize PSP in NZ9000 are well conserved in the different C-type strains (9), yet are not present in sequenced A and B type strains (12). Interestingly, the structure of the three known PSP from C-type strains is made of polymeric phospho-oligosaccharide subunits (8-10), whereas in type A (UC509.9) (13) and type B (IL1403) (14), the variable part of CWPS is constituted of short oligosaccharide chains without any repetition. These observations support the hypothesized role of the three membrane proteins in PSP polymerization. Of note, one of the identified *cwps* promoters (NZProm8, see Fig. 1) was mapped upstream of *wpsH* and *wpsG* (which are here proposed to be responsible for PSP polymerization and chain length regulation), suggesting that the genes for PSP polymerization are organized as a distinct transcriptional unit.

WpsA and WpsB are highly conserved in the different types of *L. lactis* strains (12). We propose that these two proteins initiate synthesis of the variable part of CWPS by the transfer of GlcNAc on UndP-P in all strains. Notably, in the three known type C PSP structures, GlcNAc is present in a similar position compared to the phosphodiester bond suggesting a very similar biosynthesis pathway in the three corresponding strains. Regarding IL1403, the side chain is composed of a trisaccharide that contains GlcNAc at the reducing end that could thus be synthesized in the cytoplasm on Und-P lipid, then transferred outside by a flippase (detected in the *cwps* variable part). This model can also be applied to the side chain of the type A CWPS of UC509.9 where the side chain of the rhamnose rich chains has a GlcNAc residue at the reducing end that is branched on the main chain.

In the current study, we also presented experimental evidence to elucidate the transcriptional organization of the *cwps* gene cluster in *L. lactis*. Previous work that investigated the overall transcriptome of *L. lactis* MG1363 (the parental strain of NZ9000) also highlighted potential TSS using RNAseq coupled with Terminator 50-phosphate-dependent exonuclease (TEX) treatment (35). Our promoter mapping and primer extension results match exactly the TSS for three of four promoter-containing regions (NZProm1, NZProm5, and NZProm8) identified in that study. Our results of the *L. lactis* NZ9000 *cwps* promoter mapping and primer extension analyses clearly indicate the existence of at least two distinct transcriptional units (TU_{rha} and TU_{psp}). Interestingly, a third promoter-containing region was identified which may correspond to a third transcriptional unit (TU_{pol}). The transcriptional organization aligns with previously hypothesized functional organization of the gene cluster. More specifically, TU_{rha} overlaps with the presumed rhamnan-assembly regions while TU_{psp} overlaps with that of the PSP-assembly region (11). The last transcriptional unit, TU_{pol}, encompasses *wpsH* and *wpsI* whose product are thought to control and determine total PSP polymerization, and as such may be transcriptionally control separately from the remaining cluster. The RT-PCR analysis has also highlighted transcriptional readthrough, a phenomenon which has been observed previously under certain environmental conditions in *L. lactis* (44). Finally, so far no TSS site for the NZProm7 region, which consistently showed activity both in blue/white colony X-gal assay as well as when β-galactosidase activity was measured. We hypothesize that this activity might be an artefact resulting from the cloning of the intergenic region into the pPTPL vector whereupon a previously non-existent promoter was created, leading to false-positive X-gal assay results.

Phage escape mutants obtained on a previously described PSP-deficient mutant, *L. lactis* VES5751, which carries a mutation in *llmg_0226* (*wpsH*) were still able to infect only mutants that synthesize remnant fragments of PSP. Such results give further support to our

proposed model for the biosynthesis of the PSP. Despite the infectivity of such sk1-derived phages, successful infection exhibits a minimum requirement for a trisaccharide residual PSP, such as the one exhibited by *L. lactis* NZ9000-*wpsE* (Figure 6 and Table 5), perhaps also exhibiting a Glc side chain. Residual PSP with a smaller number of sugar residues or the complete absence of PSP in the bacterial cell wall led to complete arrest of phage infectivity. This work provides further evidence to support previous suggestions on the nature of the *L. lactis* phage receptor, whereby a common trisaccharide motif found within the PSP of three C-type CWPS strains is thought to be recognized by the phage-encoded receptor binding protein (10). Interestingly, the composition of the trisaccharide motif previously identified as he *L. lactis* C-type phage receptor does not match (GlcNAc-Gal*f*-GlcNAc/Glc) the one we observed in this study (GlcNAc-Rha-GlcNAc).

We believe that the biosynthesis pathway presented here for *L. lactis* CWPS can account for the high diversity found at the level of lactococcal CWPS structures. From the different CWPS structures established for *L. lactis* strains, it is notable that the rhamnan (or rhamnose-rich) chains exhibit a certain level of variability between strains, but that the main diversity is born by the variable (PSP) chain substituents. The pathway proposed here for PSP synthesis and linkage to rhamnan can be envisioned as an evolutionary refinement of the three-component mechanisms described previously for the extracytoplasmic modification of bacterial glycoconjugates by monosaccharides (38, 43) towards a system to add complex decorative oligo/poly-saccharides onto the core rhamnan. The three component modification machinery involves (i) a glycosyltransferase able to transfer a sugar from a NDP-sugar donor to Und-P carrier, generating an Und-P-sugar, (ii) a flippase to transport the Und-P-sugar to the outer side of the cytoplasmic membrane, and (iii) a polytopic membrane protein that completes the reaction by transferring the sugar from Und-P-sugar to the glycoconjugate substrate. As an example, in *S. pyogenes*, GacI/GacJ proteins produce the GlcNAc-P-Und lipid intermediate,

which is then flipped outside the cytoplasmic membrane and GlcNAc is transferred to the polyrhamnose chain the Lancefield A carbohydrate (GAC) by GacL, a polytopic membrane protein (with 12-13 transmembrane segments) (38). Regarding *L. lactis*, this three-component mechanism is proposed to include WpsA:WpsB to synthesize GlcNAc-P-Und and to initiate synthesis of an oligosaccharide subunit, the flippase WpsG able to flip Und-P-oligosaccharide, followed by the polytopic membrane protein WpsJ to attach the PSP chain to rhamnan. In the *L. lactis* C type strains, this modification mechanism is combined with a Wzy/Wzx-like pathway to synthesize PSP chains on an Und-P lipid carrier (rather than Und-P-P) as modifying entities that are further transferred onto rhamnan chains. This patchwork/mosaic biosynthesis pathway appears to be remarkably efficient in generating diversity in CWPS structures. Further studies should aim at confirming this pathway by *in vitro* biochemical studies of enzyme activities.

4.6 References

1. Cavanagh D, Fitzgerald GF, McAuliffe O. 2015. From field to fermentation: The origins of *Lactococcus lactis* and its domestication to the dairy environment. Food Microbiology 47:45-61.
2. Bahey-El-Din M, Gahan CG, Griffin BT. 2010. *Lactococcus lactis* as a cell factory for delivery of therapeutic proteins. Current Gene Therapy 10:34-45.
3. Bahey-El-Din M, Gahan CG. 2010. *Lactococcus lactis*: from the dairy industry to antigen and therapeutic protein delivery. Discovery medicine 9:455-61.
4. Settanni L, Moschetti G. 2014. New trends in technology and identity of traditional dairy and fermented meat production processes: Preservation of typicality and hygiene. Trends in Food Science and Technology 37:51-58.
5. Price CE, Zeyniyev A, Kuipers OP, Kok J. 2012. From meadows to milk to mucosa - adaptation of *Streptococcus* and *Lactococcus* species to their nutritional environments. FEMS Microbiology Reviews 36:949-71.
6. Siezen RJ, Bayjanov JR, Felis GE, van der Sijde MR, Starrenburg M, Molenaar D, Wels M, van Hijum SAFT, van Hylckama Vlieg JET. 2011. Genome-scale diversity and niche adaptation analysis of *Lactococcus lactis* by comparative genome hybridization using multi-strain arrays. Microbial Biotechnology 4:383-402.
7. Song AAL, In LLA, Lim SHE, Rahim RA. 2017. A review on *Lactococcus lactis*: from food to factory. Microbial Cell Factories 16:15.
8. Chapot-Chartier MP, Vinogradov E, Sadovskaya I, Andre G, Mistou MY, Trieu-Cuot P, Furlan S, Bidnenko E, Courtin P, Pechoux C, Hols P, Dufrene YF, Kulakauskas S. 2010. Cell surface of *Lactococcus lactis* is covered by a protective polysaccharide pellicle. Journal of Biological Chemistry 285:10464-71.

9. Ainsworth S, Sadovskaya I, Vinogradov E, Courtin P, Guerardel Y, Mahony J, Grard T, Cambillau C, Chapot-Chartier M-P, van Sinderen D. 2014. Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. mBio 5:e00880-14.

10. Farenc C, Spinelli S, Vinogradov E, Tremblay D, Blangy S, Sadovskaya I, Moineau S, Cambillau C. 2014. Molecular insights on the recognition of a *Lactococcus lactis c*ell wall pellicle by the phage 1358 receptor binding protein. Journal of Virology 88:7005-7015.

11. Sadovskaya I, Vinogradov E, Courtin P, Armalyte J, Meyrand M, Giaouris E, Palussière S, Furlan S, Péchoux C, Ainsworth S, Mahony J, van Sinderen D, Kulakauskas S, Guérardel Y, Chapot-Chartier M-P. 2017. Another brick in the wall: a rhamnan polysaccharide trapped inside peptidoglycan of *Lactococcus lactis*. mBio 8.

12. Mahony J, Kot W, Murphy J, Ainsworth S, Neve H, Hansen LH, Heller KJ, Sørensen SJ, Hammer K, Cambillau C, Vogensen FK, van Sinderen D. 2013. Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. Applied and Environmental Microbiology 79:4385-4392.

13. Vinogradov E, Sadovskaya I, Grard T, Murphy J, Mahony J, Chapot-Chartier M-P, van Sinderen D. 2018. Structural studies of the cell wall polysaccharide from *Lactococcus lactis* UC509.9. Carbohydrate Research 461:25-31.

14. Vinogradov E, Sadovskaya I, Courtin P, Kulakauskas S, Grard T, Mahony J, van Sinderen D, Chapot-Chartier M-P. 2018. Determination of the cell wall polysaccharide and teichoic acid structures from *Lactococcus lactis* IL1403. Carbohydrate Research 462:39-44.

15. Kuipers OP, de Ruyter PGGA, Kleerebezem M, de Vos WM. 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. Journal of Biotechnology 64:15-21.

16. Oh JH, van Pijkeren JP. 2014. CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. Nucleic Acids Research 42:e131.

17. O'Driscoll J, Glynn F, Cahalane O, O'Connell-Motherway M, Fitzgerald GF, van Sinderen D. 2004. Lactococcal plasmid pNP40 encodes a novel, temperature-sensitive restriction-modification system. Applied and Environmental Microbiology 70:5546-5556.

18. Mahony J, Deveau H, Mc Grath S, Ventura M, Canchaya C, Moineau S, Fitzgerald GF, van Sinderen D. 2006. Sequence and comparative genomic analysis of lactococcal bacteriophages jj50, 712 and P008: evolutionary insights into the 936 phage species. FEMS Microbiology Letters 261:253-61.

19. Chandry PS, Moore SC, Boyce JD, Davidson BE, Hillier AJ. 1997. Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. Molecular Microbiology 26:49-64.

20. Higgins DL, Sanozky-Dawes RB, Klaenhammer TR. 1988. Restriction and modification activities from *Streptococcus lactis* ME2 are encoded by a self-transmissible plasmid, pTN20, that forms cointegrates during mobilization of lactose-fermenting ability. Journal of Bacteriology 170:3435-3442.

21. Israelsen H, Madsen SM, Vrang A, Hansen EB, Johansen E. 1995. Cloning and partial characterization of regulated promoters from *Lactococcus lactis* Tn917-lacZ integrants with the new promoter probe vector, pAK80. Applied and Environmental Microbiology 61:2540-2547.

22. Ventura M, Zink R, Fitzgerald GF, van Sinderen D. 2005. Gene Structure and Transcriptional Organization of the dnaK Operon of *Bifidobacterium breve* UCC 2003

and Application of the Operon in Bifidobacterial Tracing. Applied and Environmental Microbiology 71:487-500.

23. Kuipers OP, Beerthuyzen MM, R.J. S, W.M. V. 1993. Characterization of the nisin gene cluster nisABTCIPR of *Lactococcus lactis*. European Journal of Biochemistry 216:281-291.

24. van Pijkeren JP, Britton RA. 2012. High efficiency recombineering in lactic acid bacteria. Nucleic Acids Research 40:e76.

25. Meyrand M, Boughammoura A, Courtin P, Mezange C, Guillot A, Chapot-Chartier MP. 2007. Peptidoglycan N-acetylglucosamine deacetylation decreases autolysis in *Lactococcus lactis*. Microbiology 153:3275-85.

26. Mahony J, McGrath S, Fitzgerald GF, van Sinderen D. 2008. Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. Applied and Environmental Microbiology 74:6206-6215.

27. Lillehaug D. 1997. An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. Journal of Applied Microbiology 83:85-90.

28. Østergaard Breum S, Neve H, Heller KJ, Vogensen FK. 2007. Temperate phages TP901-1 and φLC3, belonging to the P335 species, apparently use different pathways for DNA injection in *Lactococcus lactis* subsp. *cremoris* 3107. FEMS Microbiology Letters 276:156-164.

29. Hayes S, Murphy J, Mahony J, Lugli GA, Ventura M, Noben J-P, Franz CMAP, Neve H, Nauta A, Van Sinderen D. 2017. Biocidal Inactivation of *Lactococcus lactis* bacteriophages: Efficacy and targets of commonly used sanitizers. Frontiers in Microbiology 8:107.

30. Allard STM, Giraud M-F, Whitfield C, Messner P, Naismith JH. 2000. The purification, crystallization and structural elucidation of dTDP-d-glucose 4,6-dehydratase (RmlB),

the second enzyme of the dTDP-l-rhamnose synthesis pathway from *Salmonella enterica* serovar Typhimurium. Acta Crystallographica Section D 56:222-225.

31. Blankenfeldt W, Giraud M-F, Leonard G, Rahim R, Creuzenet C, Lam JS, Naismith JH. 2000. The purification, crystallization and preliminary structural characterization of glucose-1-phosphate thymidylyltransferase (RmlA), the first enzyme of the dTDP-l-rhamnose synthesis pathway from *Pseudomonas aeruginosa*. Acta Crystallographica Section D 56:1501-1504.

32. Giraud M-F, Gordon FM, Whitfield C, Messner P, McMahon SA, Naismith JH. 1999. Purification, crystallization and preliminary structural studies of dTDP-6-deoxy-d-xylo-4-hexulose 3,5-epimerase (RmlC), the third enzyme of the dTDP-l-rhamnose synthesis pathway, from *Salmonella enterica* serovar Typhimurium. Acta Crystallographica Section D 55:706-708.

33. Marie-France G, J. MH, A. LG, Paul M, Chris W, H. NJ. 1999. Overexpression, purification, crystallization and preliminary structural study of dTDP-6-deoxy-l-lyxo-4-hexulose reductase (RmlD), the fourth enzyme of the dTDP-l-rhamnose synthesis pathway, from *Salmonella enterica* serovar Typhimurium. Acta Crystallographica Section D 55:2043-2046.

34. Rush JS, Edgar RJ, Deng P, Chen J, Zhu H, van Sorge NM, Morris AJ, Korotkov KV, Korotkova N. 2017. The molecular mechanism of N-acetylglucosamine side-chain attachment to the Lancefield group A Carbohydrate in *Streptococcus pyogenes*. The Journal of Biological Chemistry 292:19441-19457.

35. van der Meulen SB, de Jong A, Kok J. 2016. Transcriptome landscape of *Lactococcus lactis* reveals many novel RNAs including a small regulatory RNA involved in carbon uptake and metabolism. RNA Biology 13:353-366.

36. Pouwels PH, Chaillou S. 2003. Gene Expression in *Lactobacilli*, p 143-188. *In* Wood BJB, Warner PJ (ed), Genetics of Lactic Acid Bacteria doi:10.1007/978-1-4615-0191-6_6. Springer US, Boston, MA.

37. Woodward L, Naismith JH. 2016. Bacterial polysaccharide synthesis and export. Current opinion in structural biology 40:81-88.

38. Rush JS, Edgar RJ, Deng P, Chen J, Zhu H, van Sorge NM, Morris AJ, Korotkov KV, Korotkova N. 2017. The molecular mechanism of N-acetylglucosamine side-chain attachment to the Lancefield group A carbohydrate in *Streptococcus pyogenes*. Journal of Biological Chemistry 292:19441-19457.

39. Gloster TM. 2014. Advances in understanding glycosyltransferases from a structural perspective. Current Opinion in Structural Biology 28:131-141.

40. Albesa-Jové D, Giganti D, Jackson M, Alzari PM, Guerin ME. 2014. Structure-function relationships of membrane-associated GT-B glycosyltransferases. Glycobiology 24:108-124.

41. Woodward R, Yi W, Li L, Zhao G, Eguchi H, Sridhar PR, Guo H, Song JK, Motari E, Cai L, Kelleher P, Liu X, Han W, Zhang W, Ding Y, Li M, Wang PG. 2010. In vitro bacterial polysaccharide biosynthesis: defining the functions of Wzy and Wzz. Nat Chem Biol 6:418-23.

42. Reeves PR, Hobbs M, Valvano MA, Skurnik M, Whitfield C, Coplin D, Kido N, Klena J, Maskell D, Raetz CR, Rick PD. 1996. Bacterial polysaccharide synthesis and gene nomenclature. Trends Microbiol 4:495-503.

43. Mann E, Whitfield C. 2016. A widespread three-component mechanism for the periplasmic modification of bacterial glycoconjugates. Canadian Journal of Chemistry 94:883-893.

44. Dambach M, Sandoval M, Updegrove TB, Anantharaman V, Aravind L, Waters LS, Storz G. 2015. The ubiquitous yybP-ykoY riboswitch is a manganese-responsive regulatory element. Molecular Cell 57:1099-1109.

Chapter 5

Isolation and characterisation of bacteriophage escape mutants infecting PSP-deficient *Lactococcus lactis* derivatives

NB. Additional phage mutant isolation performed by the group of Dr. Marie-Pierre Chapot-Chartier, French National Institute of Agricultural Research (INRA), Jouy-en-Josas, France.

Manuscript in Preparation

Table of Contents

5.1 Abstract 212

5.2 Introduction 213

5.3 Material and Methods 215

5.3.1 Strains, phages, plasmids and growth conditions 215

5.3.2 Bacteriophage isolation and assays 217

5.3.3 Escape mutant DNA extraction and genome sequencing 218

5.3.4 Identification of SNPs needed for escape mutant phenotype 219

5.4 Results and Discussion 220

5.4.1 Isolation of sk1 escape mutants and their characterisation 220

5.4.2 Plaquing and adsorption efficiency of escape mutants 222

5.4.3 Determination of minimal SNPs needed for escape mutant phenotype 225

5.5 Conclusion 230

5.6 References 232

5.1 Abstract

The importance of the lactococcal cell-wall polysaccharide (CWPS) as a receptor for 936-type phages has previously been attested. In the current study, we describe the isolation of derivatives of 936-type phages (which we refer to as escape mutants) that acquired the ability to infect *L. lactis* strains exhibiting a polysaccharide pellicle (PSP)–deficient phenotype. Through classical phage assays, we determined that these escape mutants retain their ability to infect both the wild-type as well as PSP-deficient lactococcal strains. Interestingly, our results indicate that these mutant phages do not recognize a new type of carbohydrate, which might be more easily accessible in the PSP–deficient strains, such as the rhamnan, but instead bind to a receptor found in both versions of the strain (presumably any remaining PSP) at a significantly higher efficiency. Finally, complete phage genome sequencing of a selection of the isolated escape mutants reveals that most of the identified mutations are located within the baseplate-encoding region. Detailed mutant analysis indicates that specific amino acids within the receptor binding protein are involved in this enhanced host interaction and that the distal tail (Dit) protein apparently also contributes to this process.

5.2 Introduction

With the advent, and increased use, of defined bacterial starter cultures in industrial dairy fermentations, the issue of (bacterio)phage presence and their ability to disrupt such fermentations remains despite the implementation of numerous precautionary measures. Consequently, a deeper understanding of the intricate relationship between phages and their bacterial hosts is required in order to provide new insights and rational approaches to enhance the possibility of uninterrupted fermentations. Extensive research pertaining to *Lactococcus lactis* phage-host interactions has been undertaken and this has evolved to become a very valuable model of phage-host interactions of Gram-positive bacteria (1, 2).

Phages infecting Gram-positive bacteria recognize one or more structures or receptors on the bacterial cell wall to facilitate adsorption and infection. Such receptors may be proteins (3, 4), flagella (5, 6), or teichoic acids (7, 8). However, compelling evidence has demonstrated the role of a polysaccharidic structure on the cell wall of *L. lactis*, known as the polysaccharide pellicle (PSP), as the primary receptor for the adsorption of several groups of bacteriophages belonging to the *Siphoviridae* family (2, 9-13). As mentioned in previous chapters of this thesis, the genetic components responsible for assembling this cell wall-associated polysaccharide (CWPS) have been mapped to a well conserved region of the lactococcal genome. This gene cluster, designated *cwps*, is predicted to encode glycosyltransferases, polytopic-transmembrane proteins, a flippase, and other enzymes known to be required for polysaccharide production (10, 14, 15). Our research also identified additional genetic elements, which are located outside the *cwps* gene cluster that can decorate the CWPS by means of sugar substitutions (See Chapters 2 and 3). Finally, based on a series of mutational studies yielding *L. lactis* derivative strains with different intermediate polysaccharide structures, we also proposed a biosynthetic CWPS assembly model that consolidates currently available information regarding gene functions found within and outside the *cwps* operon (See Chapter 4 of this thesis).

Lactococcal phages have been categorised into ten distinct groups, eight of which belong to the *Siphoviridae* family, while two belong to the *Podoviridae* family (16). Of these, the most frequently isolated phages belong to the 936, P335, and c2 groups (16, 17). Phage p2 is a well-studied representative of the lytic 936 group, whose adsorption device (also known as baseplate) has been characterised at the molecular level (18, 19). This device is composed of at least three distinct structural proteins, including the distal tail (Dit) protein, the tail-associated lysin (Tal) protein, and the receptor binding protein (RBP) (18, 19). The final structure of the baseplate contains two hexamers of Dit, six trimers of RBP, and a trimer of Tal (18). A unique feature of the p2 baseplate is the conformational change it undergoes in the presence of Ca^{2+}, whereby the trimers of RBP rotate by 200^{o} from an upwards facing to a downwards (infection ready) facing orientation (18, 19). This highlights the need for a bivalent cation, such as Ca^{2+}, for this phage adsorption. Although other lactococcal phages exhibit similar baseplate modularity, they do not undergo such conformational changes, while they typically do not require cations such as calcium for host adsorption (2). Phage p2 is suggested to recognize a trisaccharide motif, GlcNAc-Gal*f*-GlcNAc-1P (or Glc-1P), found within the CWPS structure of *L. lactis* strains including its host strain MG1363 (20).

A number of *L. lactis* NZ9000 derivative strains with non-sense mutations in several genes of the *cwps* operon were constructed, all of which seem to produce no or a deficient CWPS, while exhibiting essentially complete resistance to infection by 936-type phages (See Chapter 4 of this thesis). Here, we describe the isolation of twelve so-called escape mutants that could infect PSP-deficient *L. lactis,* representing derivatives of sk1, a 936-type phage closely related to phage p2. The complete genome of eight of these escape mutants was sequenced. Finally, the ability of sk1 escape mutants to re-infect CWPS-negative lactococcal strains was mapped to mutations located within the RBP-encoding gene and, perhaps somewhat surprisingly, the Dit-encoding gene.

5.3 Material and Methods

5.3.1 Strains, phages, plasmids and growth conditions

Bacterial strains used in this study are listed in Table 1. Strains were grown at 30 ^{o}C overnight in M17 broth and/or M17 agar (Oxoid Ltd., Hampshire, United Kingdom) supplemented with glucose (5 g/L of M17 medium). Chloramphenicol (Cm^{r}) (5 μg/ml), tetracycline (Tet^{r}) (10 μg/ml) or erythromycin (Ery^{r}) (5 μg/ml) (Sigma-Aldrich, Missouri, USA) were added to the media where appropriate.

Table 1. Strains, plasmids, and bacteriophages used in this study.

Strain, plasmid, or phage	Feature(s)	Reference
Bacterial Strains		
***L. lactis* subsp. *cremoris* NZ9000**	*L. lactis* MG1363 derivative containing *nisRK*, host to phages jj50, p2, and sk1	(21)
***L. lactis* subsp. *cremoris* VES5751**	*L. lactis* MG1363 derivative exhibiting a PSP-deficient phenotype due to a spontaneous mutation in *llmg_0226*	(12)
***L. lactis* subsp. *cremoris* 3107**	*L. lactis* strains exhibiting a type C_2 CWPS genotype	(22)
Plasmids		
pNZ8048	High-copy protein expression vector, contains P^{nisA} inducible promoter, Cm^r	(23)
pNZMCC1	pNZ8048 containing a template of the RBP-encoding gene of the phage MCC1	This work
pCRISPR	High-copy vector carrying CRISPR repeats and used for integrating targeting spacer sequences, Tet^r	(24)
pCRISPR::*orf15A*	pCRISPR plasmid carrying CRISPR repeat targeting 11758-11788 region in orf15 of the phage sk1, Tet^r	This work
pCRISPR::*orf15B*	pCRISPR plasmid carrying CRISPR repeat targeting 12117-12147 region in orf15 of the phage sk1, Tet^r	This work
pCRISPR::*orf18*	pCRISPR plasmid carrying CRISPR repeat targeting 14572-14601 region in orf15 of the phage sk1, Tet^r	This work
pVPL3004	Low-copy vector expressing *cas9* along with tracRNA, Ery^r	(24)
Bacteriophages		
sk1	936 species, propagated on NZ9000	(25)
MCC1	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	Isolated by S. Kulakauskas, UCC Collection
MCC5	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	Isolated by S. Kulakauskas, UCC Collection
MCC17	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	Isolated by S. Kulakauskas, UCC Collection
IT1	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT2	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT3	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT4	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT5	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT6	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT7	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work

IT8	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work
IT9	936 species, derivative of Φsk1, propagated on *L. lactis* VES5751	This work

5.3.2 Bacteriophage isolation and assays

Phages used in this study are listed in Table 1. Propagation of phages on their respective host strains was performed as previously described (26). Similarly, both spot/plaque assays (27) and adsorption assays (28) were performed as previously described. EDTA was included in bacteriophage assays where needed to a final concentration of 50 mM. Derivatives of the 936-group phage sk1 (NC_001835.1) capable of infecting *L. lactis* VES5751 (12) were also isolated. A single plaque isolate of sk1 was propagated in 10 ml culture of *L. lactis* NZ9000, centrifuged at 5,000 rpm for 10 minutes, and the supernatant was filtered through a 0.45 μm filter, resulting in a lysate of 10^8-10^9 pfu/ml. This lysate (10 ml) was then used to perform a large scale (4 L) propagation of the phage on *L. lactis* NZ9000 and the resulting large-scale lysate was then concentrated down to 4 ml using polyethylene glycol 8000 (PEG_{8000}) precipitation as described previously (29). This resulted in a concentrated lysate of approximately 10^{11}-10^{12} pfu/ml. 10 ml of *L. lactis* NZ9000 was grown to an OD (600nm) of 0.3-0.4 at which point the culture was centrifuged and re-suspended in 1 ml of ¼ strength Ringer's (Merck, Darmstadt, Germany) supplemented with 10 mM $CaCl_2$ (Sigma-Aldrich, Missouri, USA). The concentrated sk1 lysate was added to the culture and incubated for 15 min at 30 ºC. The mixture was then centrifuged and washed three times with an equal volume of ¼ strength Ringer's solution. Finally, 250 μl of the washed cell culture was incubated with 250 μl overnight culture of *L. lactis* VES5751 in a GM17 semi-solid agar overlay. Following overnight incubation, visible plaques were picked and propagated on *L. lactis* VES5751 in order to increase the phage titre to sufficient levels to proceed with phage genome sequencing.

5.3.3 Escape mutant DNA extraction and genome sequencing

The genomic DNA of the isolated, sk1-derived escape mutants (Table 1) was extracted as previously described (30). Five escape mutants from this study (IT1-IT5), as well as the three isolates provided by S. Kulakauskas (MCC1, MCC5, MCC17), were selected for genome sequencing in order to identify the mutations allowing these escape mutants to propagate on *L. lactis* VES5751. Genome sequencing was performed using Illumina MiSeq sequencing technology (GenProbio, Parma, Italy). The MEGAnnotator pipeline was used for *de novo* sequence assemblies as well as automated gene calling (31) and putative open reading frames (ORFs) were assigned via Prodigal v2.6 and Genemark.hmm (32). A 3 kb region encompassing open reading frames (*orfs*) 15 – 18 was amplified using primers oIT147 - oIT148 (Table *2*) and sequenced (Eurofins MWG, Ebersberg, Germany) for four additional escape mutants (IT6-IT9).

Table 2. Oligonucleotides used in this study.

Oligo name	Sequence (5'-3')	Target/Comment
oIT139	*aaaaaa*ccatggcaattactgataacccag[1,2]	Fwd, 5' –end ΦMCC1 *orf18* (RBP) for cloning in pNZ8048
oIT140	*aaaaaa*tctagaatgttgcgaaaagaatcgc	Rev, 3' –end ΦMCC1 *orf18* (RBP) for cloning in pNZ8048
oIT141	*aaac*gccaaacgtttctagtttaaattctatttg*g*	Fwd, oligonucleotide for cloning CRISPR spacer into pCRISPR targeting 1st region of interest in Φsk1 - *orf15* (Dit)
oIT142	*aaaac*caaatagaatttaaactagaaacgtttggc	Rev, oligonucleotide for cloning CRISPR spacer into pCRISPR targeting 1st region of interest in Φsk1 - *orf15* (Dit)
oIT143	*aaac*ggagaaacaaatatagaacgtttaagtcgc*g*	Fwd, oligonucleotide for cloning CRISPR spacer into pCRISPR targeting 2nd region of interest in Φsk1 - *orf15* (Dit)
oIT144	*aaaac*gcgacttaaacgttctatatttgtttctcc	Rev, oligonucleotide for cloning CRISPR spacer into pCRISPR targeting 2nd region of interest in Φsk1 - *orf15* (Dit)
oIT145	*aaac*gtaatactaccatttgggtttatatcaata*g*	Fwd, oligonucleotide for cloning CRISPR spacer into pCRISPR targeting region of interest in Φsk1 – *orf18* (RBP)
oIT146	*aaaac*tattgatataaacccaaatggtagtattac	Rev, oligonucleotide for cloning CRISPR spacer into pCRISPR targeting region of interest in Φsk1 – *orf18* (RBP)
oIT147	aactagggagggcttaatgg	Fwd, 5' –end Φsk1 *orf15* (Dit) for amplification and sequencing of 3 kb region of interest
oIT148	cccatgttgcgaaaagaatc	Rev, 3' –end Φsk1 *orf18* (RBP) for amplification and sequencing of 3 kb region of interest

Footnotes:[1]Italics: Non-genomic nucleotide sequence, [2]Underlined: Restriction site, [4]Asterisk: phosphorothioate modification

5.3.4 Identification of SNPs needed for escape mutant phenotype

In order to determine the mutations in the sk1 genome that would facilitate infection of PSP-deficient *L. lactis* strains two experiments were devised. Firstly, the RBP-encoding gene of sk1 (*ORF18*) and 60-90 bp of flanking sequence was amplified from the escape mutant MCC1 to facilitate homologous recombination (using primers oIT139 – oIT140; Table *2*). A relatively short flanking region was chosen so as to avoid introduction of additional single nucleotide polymorphisms (SNPs) present in *orf16* of MCC1. The amplicon and the vector pNZ8048 were digested overnighst at 37 ^{o}C using the restriction enzymes NcoI and XbaI (Roche, Basel, Switzerland), purified using the PureLink^{TM} Quick PCR Purification Kit (Thermo Fisher Scientific, MA, USA) ligated using T4 ligase according to the manufacturer's instructions (Promega, Madison, WI, USA). The generated construct was transformed into *L. lactis* NZ9000 and was designated pNZMCC1. Cultures of *L. lactis* NZ9000 pNZMCC1 (10 ml) were grown to an OD (600nm) of approximately 0.2 and supplemented with $CaCl_2$ (Sigma-Aldrich, Missouri, USA) to a final concentration of 10 mM and 200 µl of freshly propagated wild-type sk1 (~10^9 pfu/ml). Following complete lysis of the culture, the lysate was tested against *L. lactis* VES5751. Any visible plaques were propagated on *L. lactis* VES5751 and also had the region containing the RBP-encoding gene (*orf18*) sequenced to confirm homologous recombination.

Secondly, once SNPs were identified in each of the fully sequenced escape mutants of sk1, CRISPR/Cas9-based molecular tools were employed to select for sk1 particles carrying desired mutations. Three regions in the sk1 genome, each containing one or more SNPs, were found to be of interest based on observed SNP prevalence. Two of these were located in *orf15*, which encodes the Dit protein, while one more region was found in *orf18*, predicted to encode the RBP (Figure 1). Based on this information, 30 bp CRISPR repeats targeting each of the regions were designed and cloned into the pCRISPR vector (Table 1), creating plasmids

pCRISPR::*orf15A*, pCRISPR::*orf15B*, and pCRISPR::*orf18* (Table 1). Each of these plasmids was introduced by electroporation into *L. lactis* NZ9000 which already harboured plasmid pVPL3004. Finally, sk1 wild-type was propagated until full lysis was observed on *L. lactis* NZ9000 carrying both pCRISPR and pVPL3004. The resulting lysate was used against *L. lactis* VES5751 in a plaque assay and any potential plaques were then propagated and sequenced for mutations in *orf15* or *orf18*.

5.4 Results and Discussion

5.4.1 Isolation of sk1 escape mutants and their characterisation

Previous research had identified derivative strains of *L. lactis* MG1363 and NZ9000 with a PSP-deficient phenotype caused by mutations in genes within the *cwps* operon (12) (See Chapter 4 of this thesis). Due to the apparent absence of a PSP layer, these strains are resistant to infection by a number of 936-type phages that are able to infect the wild-type strains. Through extensive screening, we isolated 12 apparently unique escape mutants of phage sk1 (Table 1) and the genomes of eight of these were fully sequenced (Table 3). Detailed comparisons between the sequenced genomes of the escape mutants with that of the parent sk1 revealed that each mutant phage had acquired between 6 and 9 SNPs. In all phages except IT4 and MCC17, the identified SNPs were located within the baseplate-encoding region indicating that the host recognition machinery had been adapted. Interestingly, the acquired mutations did not appear to lead to an extended phage host-range as various *L. lactis* strains tested (3107, IL1403, and UC509.9) remain (as they cannot be infected by the sk1 parent phage) insensitive to these escape mutants (Table 3). Finally, the phage's Ca^{2+} requirement for infection was investigated and this bivalent cation was shown to remain necessary for phage infection for all twelve escape mutants (Table 3).

Table 3. Φsk1 escape mutants sequence and phenotypic characteristics.

Phage Name	Size (bp)	Open Reading Frames	Total #SNPs	Baseplate-encoding region (*orf15-orf18*) #SNPs	$CaCl_2$ Requirement	*L. lactis* Cross-infection		
						3107	IL1403	UC509.9
sk1	28451	54	N.A.	N.A.	+	-	-	-
IT1	28452	54	5	5	+	-	-	-
IT2	28452	54	4	4	+	-	-	-
IT3	28452	54	5	5	+	-	-	-
IT4	28452	54	6	5	+	-	-	-
IT5	28452	54	7	7	+	-	-	-
MCC1	28452	54	5	5	+	-	-	-
MCC5	28452	54	5	5	+	-	-	-
MCC17	28452	54	6	5	+	-	-	-

5.4.2 Plaquing and adsorption efficiency of escape mutants

Following their isolation, the ability of the escape mutants to infect both wild-type *L. lactis* NZ9000 as well as the PSP-deficient *L. lactis* VES5751 was investigated. As demonstrated in Table 4, all escape mutants were shown to infect the wild-type strain, while also eliciting infectivity against the PSP-deficient strain, albeit at a lower efficiency (with the exception of IT1 which appears slightly more infective against *L. lactis* VES5751). However, this higher infectivity was significant only in the case of IT2, IT4, IT5, MCC1, MCC17 (p-value < 0.05), while the rest did not reach statistical significance. In order to determine any differences in the adsorption efficiency of the escape mutants, the host adsorption efficiency of four representatives of the sk1 escape mutants (IT1, IT3, MCC1, and MCC17) was investigated. The tested escape mutants exhibited a significantly higher adsorption efficiency (p-value <0.05) against both the wild-type *L. lactis* NZ9000 (with the exception of IT1 whose increased adsorption efficiency did not reach significance) as well as the PSP-deficient derivative *L. lactis* VES5751 (p-value < 0.05), when compared to the adsorption efficiency of the wild-type sk1 (Table 4). In combination, these results indicate that the sk1-derived escape mutants recognize a structure present in both wild-type and the PSP-deficient strains at a relatively higher efficiency than wild-type sk1. These results are supported by our work with a number of *L. lactis* NZ9000 *cwps* operon non-sense mutants presented in Chapter 4 of this thesis. It was shown that when certain *cwps* genes, which are thought to be involved in the later steps of the glypolymer assembly (i.e. the last two steps of subunit formation as well as polymerisation and polymer-length determination), are mutated, the *L. lactis* CWPS retained the ability to produce shortened, non-polymerised PSP. The isolated escape mutant phages are limited to infecting CWPS mutants that retained this residual PSP, yet are ineffective against strains that do not carry any PSP (fragments) on the rhamnan polymer. The *L. lactis* VES5751 cell wall

structure resembles that of *L. lactis* NZ9000-*cwpH*, as they both contain mutations in *cwpH* and both exhibit residual PSP (See Chapter 4 of this thesis).

Table 4. Phage sk1 and mutant derivatives infection and adsorption efficiencies

	Efficiency of Plaquing								
	sk1	IT1	IT2	IT3	IT4	IT5	MCC1	MCC5	MCC17
***L. lactis* NZ9000**	1	0.50 ± 0.14	1.82 ± 0.34	2.02 ± 0.44	2.09 ± 0.23	2.59 ± 0.10	3.46 ± 0.20	2.59 ± 1.17	4.06 ± 0.25
***L. lactis* VES5751**	<2.03E-08	1[a]	1[a]	1[a]	1[a]	1[a]	1[a]	1[a]	1[a]
	Relative Adsorption Efficiency								
	sk1	**IT1**	**IT2**	**IT3**	**IT4**	**IT5**	**MCC1**	**MCC5**	**MCC17**
***L. lactis* NZ9000**	1[b]	6.47 ± 0.26	NM	4.52 ± 0.74	NM	NM	7.85 ± 0.09	NM	7.90 ± 0.04
***L. lactis* VES5751**	0.16 ± 0.21	4.72 ± 0.72	NM	1.60 ± 0.70	NM	NM	6.36 ± 0.05	NM	5.07 ± 0.52

Footnote: [a] The sk1 escape mutants have as a host L. lactis VES57571, [b] The measured adsorption efficiency of sk1 against L. lactis NZ9000 was set at 1 and all other observed adsorption efficiencies presented are relative to this. NM = not measured.

5.4.3 Determination of minimal SNPs needed for escape mutant phenotype

Using the genomic sequences of each of the eight escape mutants, the predicted amino acid (aa) changes in the Dit (ORF15), Tal (ORF16) and RBP (ORF18) were compiled (Figure 1). Among the various mutants the Dit protein was predicted to be altered at six amino acids positions, with Arg186 modified in all eight mutants that were sequenced. This amino acid is placed within the "arm" domain commonly seen in Dit proteins of 936-type phages, such as p2, and which is thought to be important in the assembly of the baseplate as it interacts directly with the N-terminal domain of the RBP (19). The gene encoding the Tal protein contains the lowest number of mutations and thus amino acid alterations, present in only a subset of the escape mutants. Finally, within the gene encoding the RBP, mutations identified cause up to six amino acids to be altered. Of these amino acid alterations, mutations affecting Asn236 and/or Ile241 are observed in all assessed escape mutants and they are of particular interest due to their location within the carbohydrate-binding module (CBM) of the sk1 RBP (19, 33) and were thus considered among the most important mutations. It has previously been shown that the p2 RBP is inhibited by the phage-neutralising heavy chain antibody fragment (VHH5) obtained from llama. The two proteins have multiple points of interactions across their amino acid sequences and interestingly Asn236 was shown to form at least two hydrogen-bonds with parts of VHH5 (34, 35).

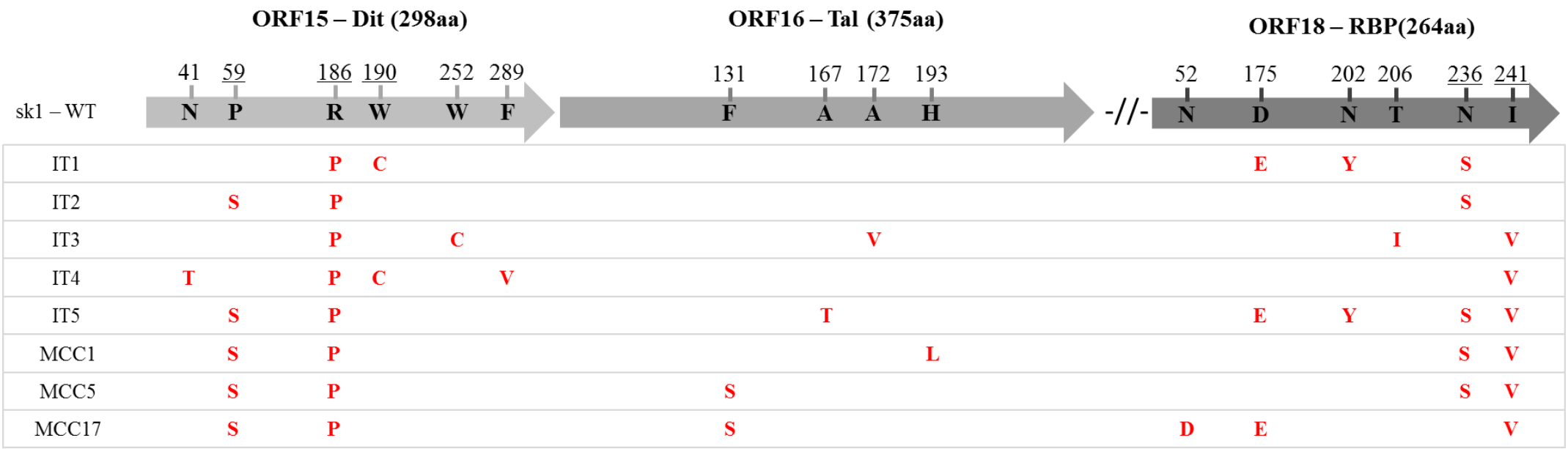


Figure 1. Schematic representation of the sk1 ORF15 putatively encoding the Dit protein, ORF16 putatively encoding the Tal protein, and finally ORF18 putatively encoding RBP. Numbers above each ORF is highlighting the location of an amino acid that was modified in the sk1 escape mutants due to a SNP. Black lettering within the ORF is indicating the amino acid present in the wild-type sk1 while red lettering underneath represents the mutated amino acid. Finally, the underlined numbers represent regions within ORFs that were targeted through CRISPR-Cas9 system to determine their importance in escape mutant phenotype.

In order to determine the essential amino acid changes required for the escape mutant phenotype, the complete RBP sequence from MCC1, which harbours only two amino acid changes (at Asn236 and Ile241), along with flanking sequence homologous to the wild-type sk1 was cloned into pNZ8048 to generate plasmid pNZMCC1. *L. lactis* NZ9000 carrying pNZMCC1 was infected with sk1 and the lysate was then tested against *L. lactis* VES5751. Similarly, *L. lactis* NZ9000 carrying an empty pNZ8048 vector was infected with sk1, as a negative control. When the sk1 lysate from *L. lactis* NZ9000 pNZMCC1 was tested against *L. lactis* VES5751 a total of 22 plaques were isolated from a lysate containing $5.70 * 10^{9}$ pfu/ml phages, indicating a rather low rate ($3.09 * 10^{-9}$) of homologous recombination. No plaques against *L. lactis* VES5751 were visible when a sk1 lysate from *L. lactis* NZ9000 pNZ8048 was tested (representing a negative control). Isolated plaques on the PSP-deficient strain were propagated and the region encompassing *orf15*-*orf18* was amplified and sequenced. All selected isolates were shown to have acquired the MCC1 RBP while carrying no mutations in *orf15* and *orf16*. This clearly indicates that the Asn236 and Ile241 changes in the RBP are the minimum requirement for the escape mutant phenotype.

In order to pinpoint individual amino acid mutations needed for the escape mutant phenotype, the CRISPR/Cas9 system was used to target regions of interest within *orf15* and *orf18* (Figure 1). The protospacer sequence, targeting the regions of interest within *orf15* and *orf18*, was cloned into the pCRISPR-targeting vectors (Table 1) and co-transformed with pVPL3004 (carrying the Cas9 endonuclease) into *L. lactis* NZ9000. Together these two vectors were intended to inhibit propagation of any sk1 particles whose regions of interest within *orf15* or *orf18* match those cloned into pCRISPR. At the same time, any sk1 particles with mutations within those regions would propagate normally and become enriched within the lysate. In this way, we had hoped to isolate sk1 derivatives containing mutations in only one of the three

regions of interest and thus be able to determine which mutations are important for conferring the escape mutant phenotype.

The strains carrying the two vectors, pCRISPR and pVPL3004 were infected with the wild-type sk1. Following each round of infection the generated lysate was mixed with wild-type sk1 and used to infect the *L. lactis* NZ9000 carrying pCRISPR and pVPL3004 until clearing of the culture was observed to indicate phage-mediated lysis. Lysis also indicates the ability of the sk1-derived phage to by-pass the CRISPR/Cas9 system through the acquisition of mutations in the protospacer and/or the protospacer-adjacent motif (PAM) sequence. In our set of experiments, sk1 lysates that by-passed the CRISPR/Cas9 system were then sequenced and shown to have acquired mutations within all protospacer and/or PAM sequences that were investigated after a maximum of four infection passages. Interestingly, when sk1 derivatives produced against pCRISPR::*orf15A* (Pro59 of the Dit) and pCRISPR::*orf18* (Asn236 or Ile241 of the RBP) were sequenced, it was determined that the mutations they had acquired did not match the mutations observed in the previously sequenced escape mutants. However, the mutations that were acquired were located within the PAM sequences and thus were sufficient for the phage to by-pass the CRISPR/Cas9 system. As such, neither set of mutations within *orf15* and *orf18* could be disregarded as important/essential mutations of the phenotype of interest. However, a single round of infection of pCRISPR::*orf15B* was sufficient to produce plaques against *L. lactis* VES5751. Once isolated, these sk1 plaques against VES5751 were sequenced and the mutations acquired matched the Arg186 mutation seen in previously sequenced escape mutants. At the same time, these mutants acquired no further SNPs in *orf16* or in *orf18*. This observation was perhaps unexpected due to the fact that the p2/sk1 Dit, despite it being an essential component of the baseplate as it directly interacts with and positions the RBP, has never been shown to be involved in phage-host recognition and/or adsorption. However, these results suggest that the Dit plays a more central role in this aspect of phage-

host interaction than previously hypothesized. As previously mentioned, this could be due to an indirect effect on the overall structure of the baseplate, as the mutation observed would be located within the region of the Dit in direct contact to the "shoulder" domain of the RBP. However, focused analysis of this particular mutation would be required to elucidate its precise *modus operandi*.

5.5 Conclusion

Our results, when taken together with the findings covered in Chapter 4, provide support to the notion that PSP is the main receptor utilized by 936-type phages including p2 and sk1. As discussed in Chapter 4, numerous attempts to acquire escape mutants of p2 and sk1 that could infect *L. lactis* derivatives that are completely devoid of PSP (such as *L. lactis* NZ9000-*cwpA* and NZ9000-*cwpB*) were unsuccessful. Such results indicate that the evolutionary leap required to recognize a completely novel carbohydrate structure, such as the rhamnan, exceeds the detection limits of our assay and would require more than a few SNPs, such as those observed in the mutants isolated in this study.

All tested sk1 escape mutants were shown to adsorb more efficiently against both the wild-type and the PSP-deficient strains. Such an ability could explain why these sk1 escape mutants are not limited by the reduced amount of PSP present on the *L. lactis* NZ9000 derivatives. Sequencing of the escape mutants and analysis of their specific SNPs support this hypothesis as all acquired mutations are located within the baseplate-encoding region of the phages. Two mutations in particular within the RBP (Asn236 and Ile241, both present within the CBD) were shown to be sufficient for the escape mutant phenotype. Interestingly, however, a previously isolated derivative of another closely related lactococcal phage known as HD09, carried the same Asn236 mutation (33). That work further illustrated the importance of the Asn236 residue for the formation of two hydrogen bonds required for interacting with a phage-inactivating antibody (33), interactions that were not exhibited by the mutant derivative of the phage. In that same study, no mutants with an altered Ile241 residue were isolated which might indicate that mutations of the Asn236 are more improrant for infection of the PSP-deficient strains. Finally, one particular amino acid change within the Dit protein of the phage (Arg186) was found in all sequenced escape mutants, highlighting for the first time, the involvement of this protein in phage-host interaction of sk1 infection, possibility through an indirect alteration

in the way the RBP units assemble and interact with the Dit protein. However, the precise mode of action of this mutation remains still unclear and further investigation into it would be required.

5.6 References

1. Dunne M, Hupfeld M, Klumpp J, Loessner M. 2018. Molecular basis of bacterial host interactions by gram-positive targeting bacteriophages. Viruses 10:397.
2. Mahony J, Cambillau C, van Sinderen D. 2017. Host recognition by lactic acid bacterial phages. FEMS Microbiology Reviews 41:S16-S26.
3. Monteville MR, Ardestani B, Geller BL. 1994. Lactococcal bacteriophages require a host cell wall carbohydrate and a plasma membrane protein for adsorption and ejection of DNA. Applied and Environmental Microbiology 60:3204-11.
4. Derkx PM, Janzen T, Sørensen KI, Christensen JE, Stuer-Lauridsen B, Johansen E. 2014. The art of strain improvement of industrial lactic acid bacteria without the use of recombinant DNA technology. Microbial Cell Factories 13:S5.
5. Raimondo LM, Lundh NP, Martinez RJ. 1968. Primary adsorption site of phage PBS1: the flagellum of *Bacillus subtilis*. Journal of Virology 2:256-264.
6. Lovett PS. 1972. PBP1: A flagella specific bacteriophage mediating transduction in *Bacillus pumilus*. Virology 47:743-752.
7. Munsch-Alatossava P, Alatossava T. 2013. The extracellular phage-host interactions involved in the bacteriophage LL-H infection of *Lactobacillus delbrueckii* ssp. *lactis* ATCC 15808. Frontiers in Microbiology 4:1-5.
8. Li X, Gerlach D, Du X, Larsen J, Stegger M, Kühner P, Peschel A, Xia G, Winstel V. 2015. An accessory wall teichoic acid glycosyltransferase protects *Staphylococcus aureus* from the lytic activity of Podoviridae. Scientific Reports 5:17219.
9. Ainsworth S, Sadovskaya I, Vinogradov E, Courtin P, Guerardel Y, Mahony J, Grard T, Cambillau C, Chapot-Chartier M-P, van Sinderen D. 2014. Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. mBio 5:e00880-14.

10. Mahony J, Kot W, Murphy J, Ainsworth S, Neve H, Hansen LH, Heller KJ, Sørensen SJ, Hammer K, Cambillau C, Vogensen FK, van Sinderen D. 2013. Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. Applied and Environmental Microbiology 79:4385-4392.

11. Mahony J, Randazzo W, Neve H, Settanni L, van Sinderen D. 2015. Lactococcal 949 group phages recognize a carbohydrate receptor on the host cell surface. Applied and Environmental Microbiology 81:3299-305.

12. Chapot-Chartier MP, Vinogradov E, Sadovskaya I, Andre G, Mistou MY, Trieu-Cuot P, Furlan S, Bidnenko E, Courtin P, Pechoux C, Hols P, Dufrene YF, Kulakauskas S. 2010. Cell surface of *Lactococcus lactis* is covered by a protective polysaccharide pellicle. Journal of Biological Chemistry 285:10464-71.

13. Farenc C, Spinelli S, Vinogradov E, Tremblay D, Blangy S, Sadovskaya I, Moineau S, Cambillau C. 2014. Molecular insights on the recognition of a *Lactococcus lactis c*ell wall pellicle by the phage 1358 receptor binding protein. Journal of Virology 88:7005-7015.

14. Dupont K, Vogensen FK, Neve H, Bresciani J, Josephsen J. 2004. Identification of the Receptor-Binding Protein in 936-Species Lactococcal Bacteriophages. Applied and Environmental Microbiology 70:5818-5824.

15. Sadovskaya I, Vinogradov E, Courtin P, Armalyte J, Meyrand M, Giaouris E, Palussière S, Furlan S, Péchoux C, Ainsworth S, Mahony J, van Sinderen D, Kulakauskas S, Guérardel Y, Chapot-Chartier M-P. 2017. Another brick in the wall: a rhamnan polysaccharide trapped inside peptidoglycan of *Lactococcus lactis*. mBio 8.

16. Deveau H, Labrie SJ, Chopin M-C, Moineau S. 2006. Biodiversity and classification of lactococcal phages. Applied and Environmental Microbiology 72:4338-4346.

17. Mahony J, van Sinderen D. 2014. Current taxonomy of phages infecting lactic acid bacteria. Frontiers in Microbiology 5:7.

18. Bebeacua C, Tremblay D, Farenc C, Chapot-Chartier M-P, Sadovskaya I, van Heel M, Veesler D, Moineau S, Cambillau C. 2013. Structure, Adsorption to Host, and Infection Mechanism of Virulent Lactococcal Phage p2. Journal of Virology 87:12302-12312.

19. Sciara G, Bebeacua C, Bron P, Tremblay D, Ortiz-Lombardia M, Lichière J, van Heel M, Campanacci V, Moineau S, Cambillau C. 2010. Structure of lactococcal phage p2 baseplate and its mechanism of activation. PNAS 107:6852-6857.

20. McCabe O, Spinelli S, Farenc C, Labbé M, Tremblay D, Blangy S, Oscarson S, Moineau S, Cambillau C. 2015. The targeted recognition of *Lactococcus lactis* phages to their polysaccharide receptors. Molecular Microbiology 96:875-886.

21. Kuipers OP, de Ruyter PGGA, Kleerebezem M, de Vos WM. 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. Journal of Biotechnology 64:15-21.

22. Braun V, Hertwig S, Neve H, Geis A, Teuber M. 1989. Taxonomic differentiation of bacteriophages of *Lactococcus lactis* by electron microscopy, DNA-DNA hybridization, and protein profiles. Microbiology 135:2551-2560.

23. McGrath S, Fitzgerald GF, van Sinderen D. 2001. Improvement and optimization of two engineered phage resistance mechanisms in *Lactococcus lactis*. Applied and Environmental Microbiology 67:608-16.

24. Oh JH, van Pijkeren JP. 2014. CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. Nucleic Acids Research 42:e131.

25. Chandry PS, Moore SC, Boyce JD, Davidson BE, Hillier AJ. 1997. Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. Molecular Microbiology 26:49-64.

26. Mahony J, McGrath S, Fitzgerald GF, van Sinderen D. 2008. Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. Applied and Environmental Microbiology 74:6206-6215.

27. Lillehaug D. 1997. An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. Journal of Applied Microbiology 83:85-90.

28. Østergaard Breum S, Neve H, Heller KJ, Vogensen FK. 2007. Temperate phages TP901-1 and φLC3, belonging to the P335 species, apparently use different pathways for DNA injection in *Lactococcus lactis* subsp. *cremoris* 3107. FEMS Microbiology Letters 276:156-164.

29. Hayes S, Murphy J, Mahony J, Lugli GA, Ventura M, Noben J-P, Franz CMAP, Neve H, Nauta A, Van Sinderen D. 2017. Biocidal Inactivation of *Lactococcus lactis* bacteriophages: Efficacy and targets of commonly used sanitizers. Frontiers in Microbiology 8:107.

30. Lavelle K, Martinez I, Neve H, Lugli GA, Franz CMAP, Ventura M, Bello FD, Sinderen Dv, Mahony J. 2018. Biodiversity of *Streptococcus thermophilus* phages in global dairy fermentations. Viruses 10:577.

31. Lugli GA, Milani C, Mancabelli L, van Sinderen D, Ventura M. 2016. MEGAnnotator: a user-friendly pipeline for microbial genomes assembly and annotation. FEMS Microbiology Letters 363.

32. Besemer J, Borodovsky M. 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Research 27:3911-20.

33. Tremblay DM, Tegoni M, Spinelli S, Campanacci V, Blangy S, Huyghe C, Desmyter A, Labrie S, Moineau S, Cambillau C. 2006. Receptor-binding protein of *Lactococcus lactis* phages: identification and characterization of the saccharide receptor-binding site. Journal of Bacteriology 188:2400-10.

34. Spinelli S, Desmyter A, Verrips CT, de Haard HJW, Moineau S, Cambillau C. 2005. Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nature Structural & Molecular Biology 13:85.

35. Tremblay DM, Tegoni M, Spinelli S, Campanacci V, Blangy S, Huyghe C, Desmyter A, Labrie S, Moineau S, Cambillau C. 2006. Receptor-binding protein of *Lactococcus lactis p*hages: identification and characterization of the saccharide receptor-binding site. Journal of Bacteriology 188:2400-2410.

Chapter 6

Discussion

The economic importance of *Lactococcus lactis* for industrial fermentations is widely documented. As a consequence, the interactions between lactococci and their infecting phages has been subject to intense research scrutiny over the past decades. Although this phage/host relationship has been extensively studied from the phage perspective with notable advances in structure-function analyses of phage-encoded adhesion devices, work on lactococcal host receptor structures and their biosynthesis is still in its infancy. Seminal work on the lactococcal pellicle (PSP), as part of the cell wall polysaccharide (CWPS), has determined both the structural diversity exhibited by a number of different *L. lactis* strains, as well as the role that this variable glycan plays in determining phage/host interactions. The aim of this thesis was to further expand our knowledge on the molecular components involved in its biosynthesis, to determine the functionality of the CWPS-assembly enzymes, as well as deciphering the extent to which such structural differentiation can impact on cell survivability and physiology, and/or phage resistance. Finally, this thesis reports on the identification of various substrate-specific three-component glycosylation systems of three different cell wall glycopolymers in *L. lactis*. Furthermore, an expanded model for the biosynthetic pathway involved in the assembly of lactococcal CWPS is proposed.

In Chapter II, the identification and characterisation of the genes responsible for the glycosylation of the *L. lactis* PSP was reported. The genes shown to be involved in PSP glycosylation were termed *csdC*, encoding a presumed C_{55}-P GT, *csdD*, encoding a PolM GT, and *cflA*, encoding a flippase. The first two genes, *csdC* and *csdD*, are present in tandem in the lactococcal genome, whereas *cflA* is located elsewhere on the chromosome. CsdC is believed to initiate the process of glycosylation through the transfer of a UDP-glucose onto an undecaprenyl phosphate carrier in the cell membrane, which is then re-oriented (or "flipped") to the outer side of the cytoplasmic membrane through the activity of CflA. The CsdD protein is then presumed to catalyse the attachment of the glucose moiety on a specific position of the

PSP subunit. Of the three identified enzymes, CsdD's function was shown to be critical for the process of PSP glycosylation whereas eliminating the function of CsdC or CflA by genetic mutation reduced but did not completely eliminate the presence of PSP-specific glucose substitutions. When the function of CsdC is inhibited, it can apparently be complemented by the function of another C_{55}-P GT, CsdA, which was also identified in this study. Similarly, the presence of residual glucose side-chains on PSP in the *cflA*-negative mutant indicates that another, yet unidentified, flippase remains functional and works in parallel with CflA. Our inability to identify another flippase based on amino acid sequence similarity to CflA may be explained by the presence of transmembrane domains in the encoded enzyme. Such domains tend to exhibit a high degree of sequence diversity even between enzymes of similar function due to the common presence of several hydrophobic domains in their encoded sequence. The presence of this set of genes explains the apparent disparity between the CWPS structure of *L. lactis* MG1363 and *L. lactis* 3107 and the number of glycosyltransferase-encoding genes present in the *cwps* gene cluster of the genomes of these strains. Finally, similar to previous evidence of phage receptor glycosylation impacting phage infectivity, this thesis presents further evidence to support that such systems may be maintained in lactococcal genomes as a phage defence mechanism. In fact, during the investigation into the distribution of such genes in different *L. lactis* strains, homologues of *csdCD* were located within predicted prophage gene clusters, perhaps indicating a phage origin of these genes. Interestingly, the glycosylation of the lipopolysaccharide of *Salmonella* and *Shigella* species has been linked to prophage-encoded three-component glycosylation systems. In these cases, changes in glycosylation patterns result in serotype conversions of the respective strains. Preliminary evidence suggests that similar glycosylation systems are also exploited by Gram-positive genera such as *Bacillus, Listeria,* and *Streptococcus* for the sugar substitutions found on teichoic acid structures.

In addition to the set of genes mentioned above, Chapter III describes the characterisation of four additional genes, which are genetically organised in two pairs, *csdAB* and *csdEF*, being involved in rhamnan glucosylation and lipoteichoic acid galactosylation, respectively. As observed for the glucosylation of PSP, the glucosylation of rhamnan appears to confer a degree of phage insensitivity to the host. In a subset of the phages tested, this phenotype was shown to be mediated by adsorption inhibition. However, this reduction in phage infectivity was not always linked to adsorption interference, but rather to subsequent infection steps, such as DNA injection or phage particle release, thus requiring further study into the exact nature of this phenomenon. LTA galactosylation does not appear to affect phage infection, at least not for the phages employed in this study, and therefore we could not unearth any evidence for LTA as a lactococcal phage receptor. Consequently, the galactose substitutions of this glycopolymer may affect other aspects of bacterial physiology. Although preliminary evidence showed links between LTA galactosylation and bacteriocin resistance, future studies should further explore the importance of such modifications. Interestingly, previous work evaluating the sequence diversity of the *cwps* gene cluster between *L. lactis* strains, highlighted the potential of a strain genotyping mechanism based on such sequence differences for starter culture strain selection. However, in the future such variably present glycosylation systems in *L. lactis*, like the three presented in this thesis, should also be taken into account when such genotyping schemes are being established. Perhaps, we are faced with a situation similar to what has been observed for the serotyping efforts of streptococcal species where classical differentiation techniques based on a culture's reaction to antisera failed to accurately decipher a strain's CWPS. Genotyping that also includes the genes encoding known glycosyltransferase systems presents a much more precise and definitive strategy. One final limitation of such a differentiation scheme stems from the fact that, as we have already observed, presence of a gene does not always guarantee the functionality of its encoded protein.

For example, there were cases where due to transposonal elements in the genes encoding for the glycosyltransferase systems we identified, the structural modifications these systems imparted were not observed in the strain's final glycopolymer structure. Similarly, as in the case of *csdAB* in *L. lactis* NZ9000, despite the presence and full functionality of both genes, their glycosylation effect on the rhamnan structure was minimal due to their extremely low levels of transcription. As a result of this, more focus would need to be placed on identifying the molecular and environmental cues that control the expression of such glycosylation systems. Such effects must also be taken into consideration when implementing any genotyping systems based on *cwps* and glycosylation genes.

The transcriptional organisation as well as deduced functionalities of the genes within the *cwps* gene cluster were presented in Chapter IV. Through extensive promoter mapping as well as primer extension analysis, two active promoter regions were identified within the gene cluster. These regions matched a previously speculated division of the cluster into a rhamnan-assembly operon and PSP-assembly operon. A third promoter region with rather low activity was found within the 3'-end of the PSP-assembly operon upstream of *cwpH* and *cwpI*. The products of these latter genes are believed to be involved in the polymerisation and chain-length determination of the PSP, respectively, as presented in our model (Chapter 4). The model for the CWPS assembly pathway presented here provides a framework to further investigate the functionalities of the individual glycosyltransferases and polysaccharide-assembly enzymes encoded by the *cwps* gene cluster. The specificities of the enzymes presented here were based on structural analysis of the CWPS of mutated lactococcal strains, however, future efforts may need to focus on the expression, purification, and functional characterisation of these proteins. Conclusive functional assignments would be essential for any efforts in predicting CWPS structures of novel lactococcal strains based solely on genomic data. Additionally, with the CWPS structure of representatives from both *cwps* A and B genotypes being available,

similarities and differences in gene functionalities between the different genotypes would need to be investigated. In the future the use of this proposed model for CWPS assembly would provide an invaluable standard to which newly isolated *L. lactis* strains could be compared when trying to structurally define their potentially novel CWPS structures. Finally, knowledge of the role each of these *cwps* genes play in determining the final CWPS structure would help us further improve upon the specifity of the existing *cwps* genotyping scheme. In other words, future primer design should be focused on those genes with distinct functionalities between the various existing *cwps* genotypes as defined by the model proposed here.

The association between the rhamnan and the PSP in *L. lactis* strains belonging to the *cwps* C genotype was investigated further in Chapter IV. The presence of a covalent link between the two glycopolymers was previously hypothesised though direct evidence for this notion has yet to be provided. With the data presented here, it has become apparent that the rhamnan forms the "backbone" structure of the lactococcal CWPS, similar to the RhaCWPS exhibited by streptococcal strains, while the PSP forms antenna-like substitutions attached onto the rhamnan. Isolation of the two glycopolymers as separate glycans is thought to be an artefact of the polysaccharide extraction methods used in the analysis. Altogether, *L. lactis* appears to employ three different types of polysaccharide-production systems. Firstly, production of the rhamnan employs an ABC transporter-dependent system whereby the whole molecule is polymerised and assembled intracellularly, transported across the membrane, and then attached onto the peptidoglycan layer. Secondly, production of the PSP employs a Wzy-dependent system whereby the subunits of the PSP are assembled intracellularly. A Wzx flippase then transports the subunit to the periplasmic area where they are polymerised and attached to the rhamnan. Finally, two distinct membrane-associated three component glycosylation systems are used to attach monosaccharide substitutions to the rhamnan and the PSP.

The importance of the PSP as a phage receptor was further explored in Chapter V. A number of so-called escape mutants were isolated that appear to have an increased affinity for the PSP component of the CWPS. Such a characteristic allows these phages to infect certain strains with a PSP-deficient phenotype. The isolation of the phages with increased adsorption efficiencies against the PSP-carrying strains could signal a potentially beneficial industrial application of their RBPs. More specifically, these highly efficient RBPs could be heterogeneously expressed, purified, and subsequently introduced in industrial fermentations where their presence could potentially block further infection from naturally present lytic phages. Further study into the use of these increased affinity RBPs as potential phage infection competitors would certainly be of interest.

Overall, this thesis significantly expands on previous work done concerning the *L. lactis* CWPS structure and functionality in relationship to phage/host interactions. Analysis of novel molecular pathways involved in CWPS and LTA structural determination has resulted in significant advancements in our understanding of the glycopolymer production by and cell wall structure of *L. lactis*. As we expand this knowledge to incorporate additional lactococcal strains of divergent *cwps* genotypes and CWPS structures, the maintenance and importance of such glycopolymer assembly systems across the whole species should be investigated. Based on the expected findings interactions between and classification based on known phage anti-receptors and their host receptors will be further detailed. Finally, knowledge on the receptor assembly pathways should allow for a better prediction of these phage receptors, which may consequently allow for the development of more bacteriophage-robust starter strains.

Acknowledgements

Firstly, I would give my deepest thanks to Douwe and Jenn for giving me the chance to work on this PhD and welcoming me into this amazing group of people with such great sense of family and community! Your presence, support, and guidance was truly invaluable and I am truly grateful for having done my PhD in such an environment. Jenn, I have no clue how you do it, but a chat with you makes all the science-blues and worries go away! Thanks for all the times that you put my mind at ease and gave me fresh confidence and inspiration to continue. ☺

To all our amazing collabolators, Marie-Pierre Chapot-Chartier, Irina Sadovskaya, and Pascal Courtin, a great big thank you for delivering such amazing work and putting up with my constant email pestering for the latest data!

To the countless people that I worked with in our labs 339/340, Joana, Katherine, Stephen, Andrea, Tom, Eoghan, Phil, James, Brian, Juan, and so many others. Thank you for being great people to work with, thanks for putting up with my wild mood swings and weird personality, especially in those middle-dark years, and thank you for being great friends! Joana, being there for little Afonso and your family in the last few months has given me some amazing times to remember! I hope we'll all have a lot more of these to come, whether it be here, Portugal, or Germany! Your advice on life and work, I will always cherish.

Το my family in Greece! Επιτέλους τέλειωσε! Ο αιώνιος φοιτητής αποφοιτά! Τα τελευταία δύο χρόνια δεν ήταν και τα πιο εύκολα για την οικογένεια μας και μέσα σε όλα αυτά έπρεπε να βγάλω και ένα διδακτορικό… αλλά με τη βοήθεια σας, τις συμβουλές σας, την αγάπη σας τελειώνει ακόμη ένα μεγάλο κεφάλαιο. Ελπίζω τώρα που θα με πιο κοντά θα έρχομαι να σας

βλέπω ακόμη πιο συχνά! Ειδικά όλα τα μικρά μου ανίψια. Στο τέλειο μπαμπά μου και στην μοναδική υπέρτατη πολυαγαπημένη μου Ντένια, σας αγαπώ, σας λατρεύω, και σας ευχαριστώ που είστε δίπλα μου. Τα λόγια δεν μπορούν να εκφράσουν πόσο πολύ με βοηθήσατε!

To Marine, my love, partner, my companion, my support, my everything. I wouldn't be writing any of this if you hadn't been there. You've helped me through the most difficult time of my life and you did it selflessly. You listen to me when I ramble on and on and on about the most inane things, you dream with me, you get excited with me, and you are just amazing. Thank you for being a part of my life and making it 10logs better! This is the end of my PhD chapter and yours is coming up soon, I can't wait to be there for you when you need me!

Finally, to the person, without whom I wouldn't be who I am today, my amazing mother. Σε ευχαριστώ που δεν έπαψες ποτέ να με πιέζεις, να πιστεύεις στις δυνατότητες μου, να με ενθαρύνεις. Δεν θα το ξεχάσω ποτέ. Κάθε μέρα σε σκέφτομαι και μου λείπεις. Ότι έχω πετύχει είναι εξαιτίας σου και ελπίζω να με βλέπεις και να χαμογελάς με το μοναδικό σου χαμόγελω! Σ'αγαπώ!

You are all grand! Thanks!